SONG.LY

MUSIC GENRE AND SENTIMENT ANALYSER

A13011: MACHINE LEARNING AND PATTERN RECOGNITION

CONTENTS

PROBLEM STATEMENT

CONTEXT

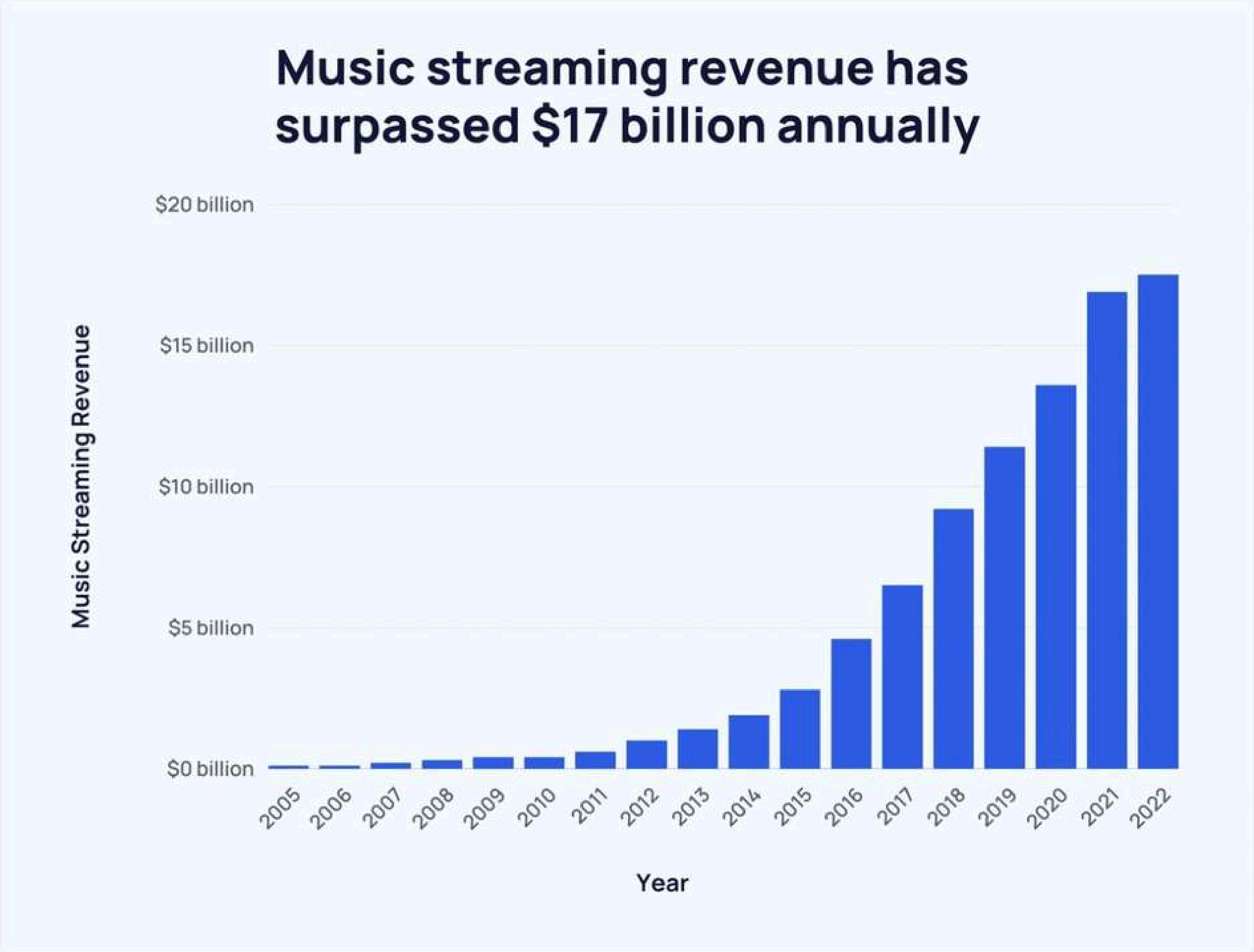
Music Streaming Services (Top Stats)

- Music streaming makes up **84%** of music industry revenue
- The music streaming industry grew by **over 10%** over the last year
- Music streaming's global revenue currently sits at **$17.5 billion**
- Paid music streaming makes up **23%** of all music streaming
- **78%** of people listen to music via a streaming service
- **Over 600 million** subscribe to a music streaming platform

Music Streaming Services Stats (2024)

Between 2010 and 2020, revenue increased by **around 34x** from **$0.4 billion** to **$13.6 billion.**

And in 2022, music streaming revenue stood at approximately **$17.5 billion**.



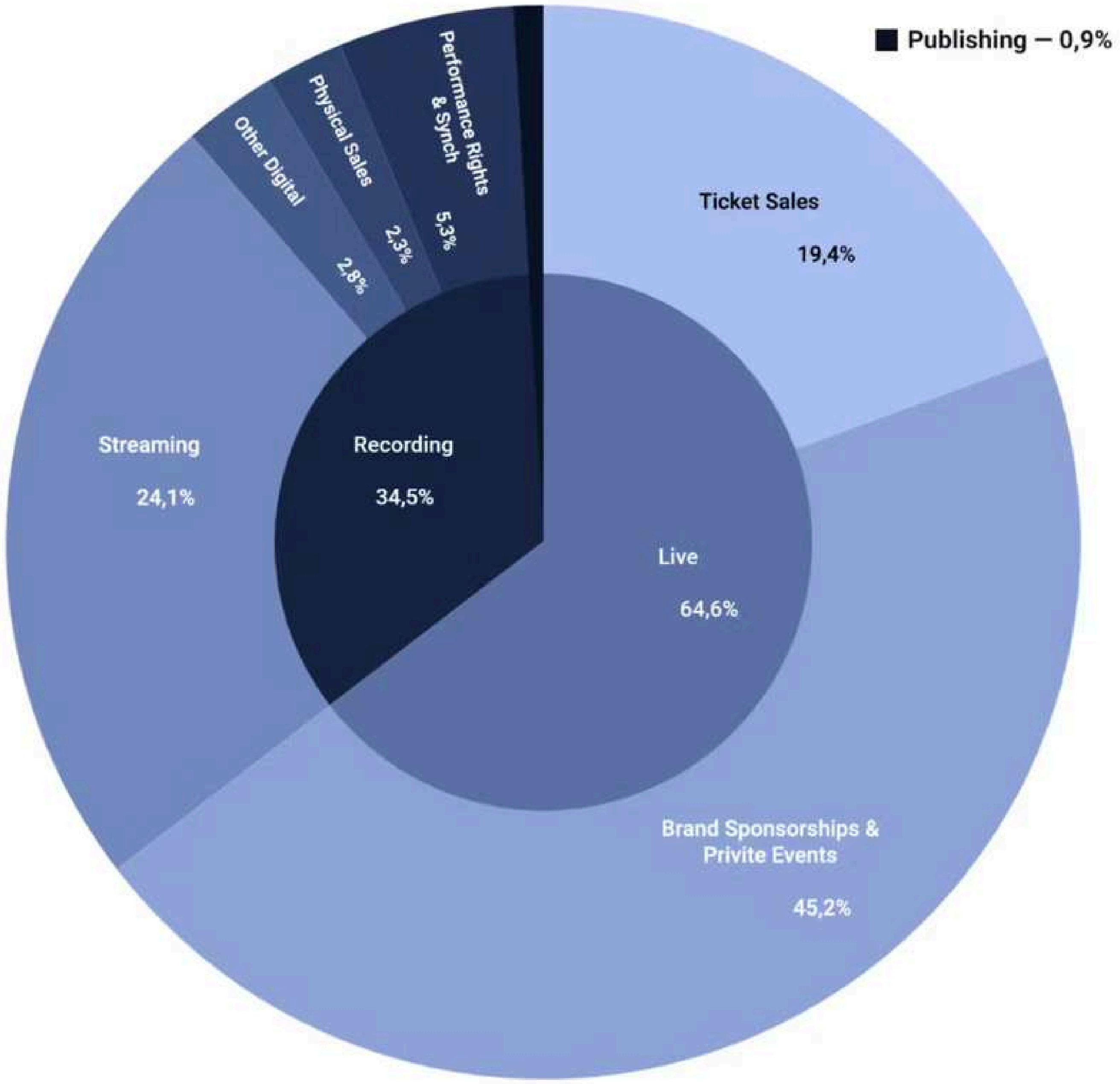
Music Consumption Source	Percentage
Paid music streaming	23%
Video streaming	22%
Radio	16%
Short videos (TikToks)	11%
Ad-supported music streaming	9%
Purchased music (CDs, downloads)	9%
Other (Netflix, music borrowing)	5%
Social media	3%
Live shows	2%

Around 4 in 5 people listen to music using a streaming service (IFPI)

Approximately **78%** listen to music using some form of music streaming service.

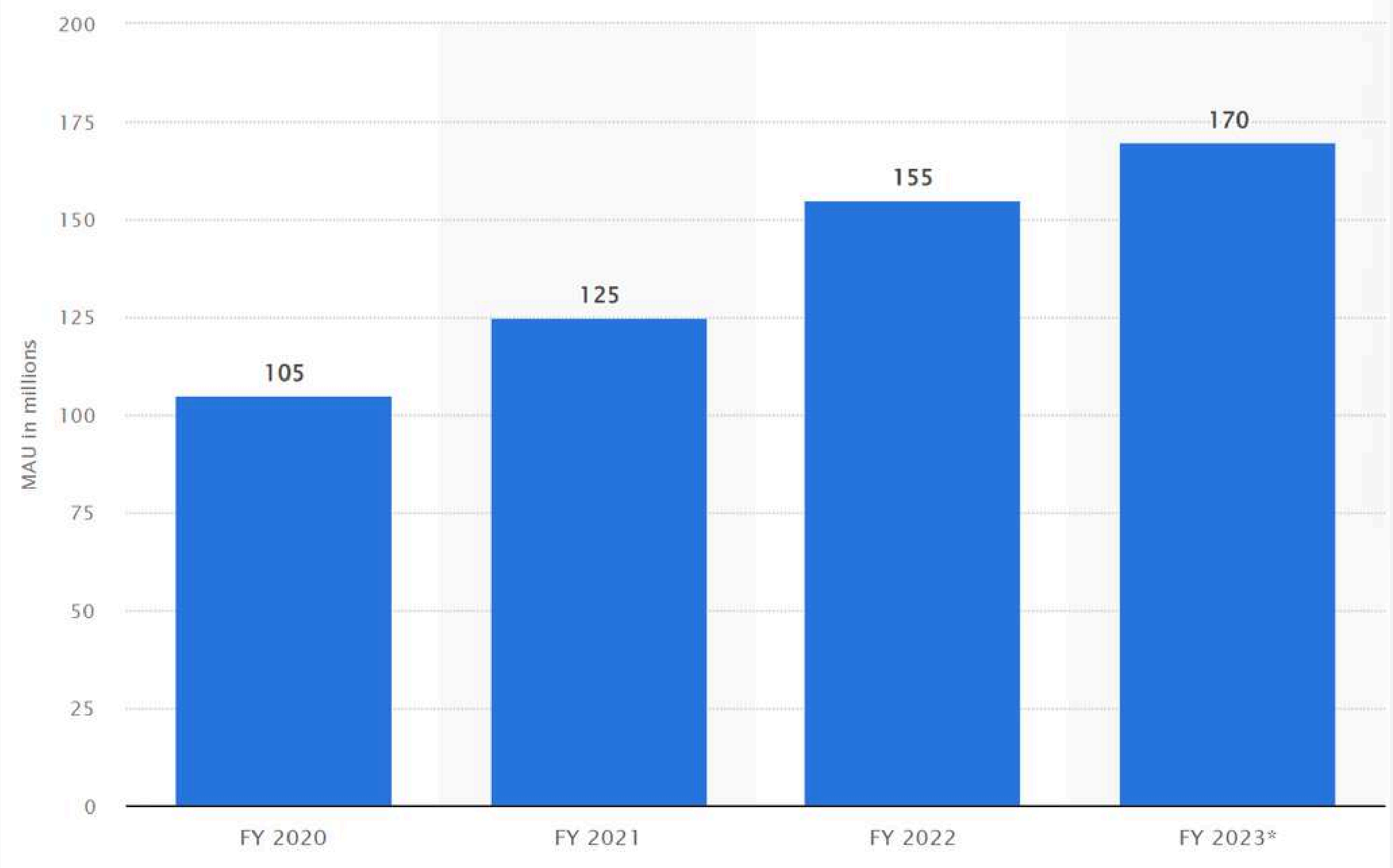
Music Streaming Services Stats (2024)

Summing up the revenues across the three main sub-industries, we estimate the scope of the Indian music market at **$443 million**.



Indian Music Industry Revenue by Source, 2018
Sources: IFPI, PwC, VISION 2022, IPRS

Average monthly active users of music streaming services in India from fiscal year 2020 to 2023 (in millions)



Statista 2024

Indian Music Industry Analysis: Streaming, Live Industry, Bollywood, 2022 Trends, and More

PROBLEM STATEMENT

In today's digital age, music plays a vital role in expressing our emotions. Yet, understanding the feelings and genres in songs in this ever-evolving industry can be tricky. **Our Model identifies** the most likely **genre and sentiment** the song is trying to convey. We do this over **10 genres (including bollywood)** and **6 sentiments.**
Our goal is to help people better understand themselves and enjoy and appreciate the music they listen to more.

POTENTIAL APPLICATIONS

Emotional Analysis:

- Our model can help users analyze the emotions conveyed by songs, providing insights into their mood and feelings through their music preferences, as well as helping them understand what type of music they enjoy listening to.

Educational purposes:

- Our solution can be utilised by individuals with difficulties in processing emotions/music students to learn about emotions and genres, offering a practical application for understanding these music attributes for personal enhancement and awareness.

Scope in Industry (for industries working with music):

- Our Model can be used as a module for precise music analysis, which would allow one to enhance user experiences, tailor recommendations, and drive engagement of apps too!

IMPACT

Deeper Music Appreciation: It helps users recognize patterns in their listening habits and discover new music that matches their mood.

Improved Emotional Awareness: The model aids users in understanding their emotional state through music preferences. It contributes to self-reflection and improved well-being.

Music-Based Icebreakers: Feeling awkward at a social gathering? Our model can be used to analyze a song everyone knows and discuss the emotions or genre it conveys.

LITERATURE REVIEW

PAPER 1:

MUSICAL EMOTIONS ANALYSIS

(TISMO-CAPILI, 2020-21)

This project's aim is to study and compare machine learning techniques that can help identify the emotions of a person by first finding the emotions that are conveyed through the music they listen to.

Dataset:
A new dataset was created by extracting songs from Spotify playlists representing various (8) emotions, ensuring balance in the number of songs per mood. Songs were retrieved and their audio features were processed and merged into a single dataset.

The models that have LOOCV implemented will mostly be discussed here as they result, they show are the most representative of the dataset. The scoring that will mostly be focused on will be the custom scoring method that was created because the score that is generated by this is more representative to the models than the normal scoring method. The score produced by this method will be referred to as the 'Lenient Accuracy' score.

Reference: T. Tismo-Capili, "Musical Emotions Analysis," thesis, 2021.

	name	artists	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signature	valence	nlp_lyrics	nlp_annotations	valence+nlp	mood
0	Remember Me	[UMI]	0.45600	0.840	199227	0.344	0.000034	5	0.3500	-8.613	0	0.0374	111.994	4	0.526	0.7286	0.7286	0.5406	Chill
1	North Face	[ODIE]	0.79200	0.802	196800	0.382	0.163000	10	0.0783	-7.356	1	0.0312	99.969	4	0.581	0.9486	0.9486	0.6000	Chill
2	Mine	[Alex Isley, Jack Dine]	0.82800	0.347	212571	0.395	0.000011	6	0.1250	-9.278	0	0.0567	67.492	4	0.133	0.8880	0.8880	0.1508	Chill
3	Shine	[Cleo Sol]	0.68600	0.742	226118	0.504	0.517000	1	0.1030	-10.105	1	0.0392	140.000	4	0.601	0.9515	0.9515	0.6200	Chill
4	Loverboy	[Joesef]	0.59400	0.356	238123	0.611	0.000000	11	0.1190	-7.219	0	0.0567	79.338	4	0.524	0.9765	0.9765	0.5435	Chill
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
95	Blind Ma	[Xavier Omär]	0.45300	0.890	242196	0.523	0.000002	6	0.0831	-8.526	1	0.0560	111.031	4	0.345	0.9992	0.9992	0.3650	Chill
96	Peaches (feat. Daniel Caesar & Giveon)	[Justin Bieber, Daniel Caesar, Giveon]	0.32100	0.677	198082	0.696	0.000000	0	0.4200	-6.181	1	0.1190	90.030	4	0.464	0.9889	0.9889	0.4838	Chill
97	So Good At Being in Trouble	[Unknown Mortal Orchestra]	0.03630	0.829	230147	0.435	0.878000	0	0.1190	-10.136	1	0.0515	103.816	4	0.594	-0.7326	-0.7326	0.5793	Chill
98	Somehow.	[Phony Ppl]	0.71900	0.502	230973	0.399	0.003630	11	0.6450	-9.934	1	0.0295	92.904	4	0.124	0.2887	0.2887	0.1298	Chill
99	Missing Out	[Syd]	0.00761	0.705	239744	0.528	0.004120	11	0.1290	-5.582	1	0.0416	119.816	4	0.272	-0.9540	-0.9540	0.2529	Chill

100 rows × 19 columns

Figure 8 Example of what the extracted song data looks like in a Pandas data frame

The evaluated models had Leave-One-Out Cross-Validation (LOOCV) implemented. Results indicated SVM as the most accurate model, outperforming Gaussian Naïve Bayes by a margin of 2%.

Classifier	Lenient Accuracy	%
Decision Tree	0.7491782553729498	74.91%
SVM	0.816586592414686	81.65%
K-Nearest Neighbours	0.7173198482932994	71.73%
Gaussian Naïve Bayes’	0.792844500632114	79.28%

- Can build dataset without any Ethical Concerns
- Minimal Bias Sentiment Representation
- Limited Dataset Size

PAPER 2:

MUSIC GENRE CLASSIFICATION USING RANDOM FOREST

PANDITA S., 2021

Music Genre Classification using Random Forest

Sidharth Pandita · Follow
Published in hackerdawn · 5 min read · May 29, 2021

This project is using Random Forest for music genre classification (10) using music audios of 30 seconds each.

Kaggle Dataset: GTZAN Dataset - Music Genre Classification (see later)

Model Creation & Prediction

It's time to create our model. We will use Random Forest Classifier to built the model. We'll fit the model using the training data and predict the testing data. Our model's accuracy turns out to be 81.38 %, which is great!

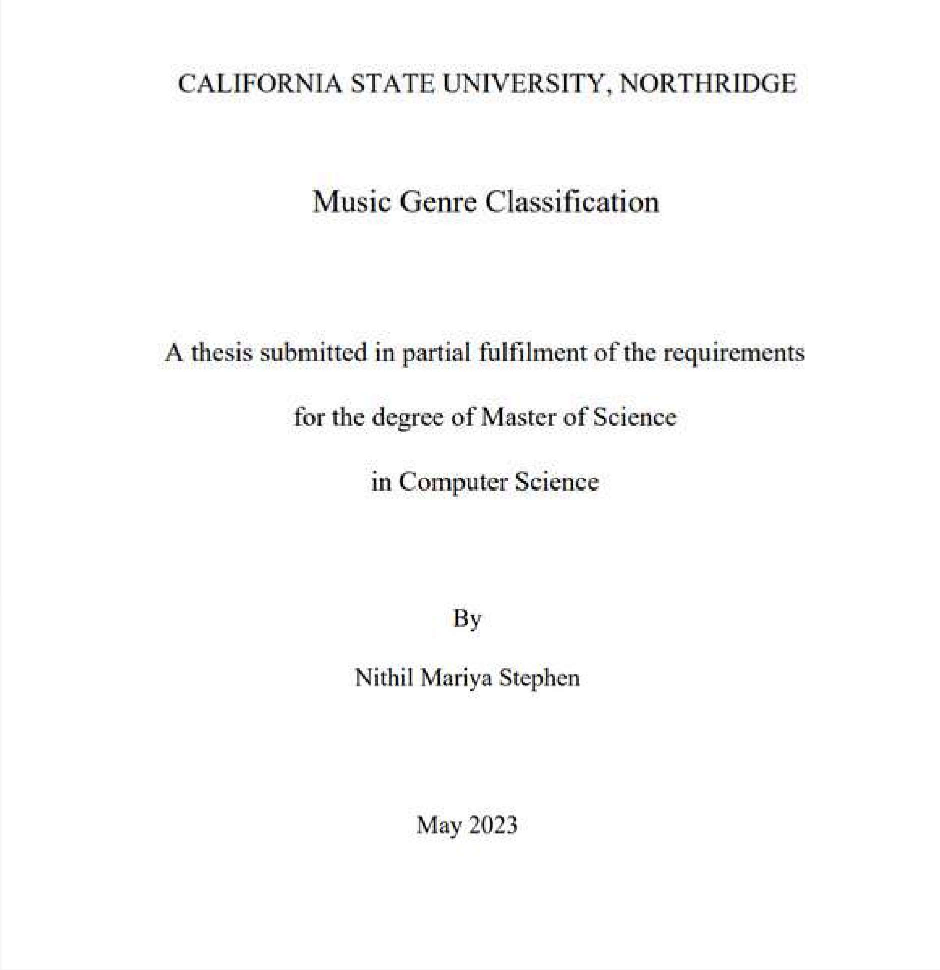
Reference: S. Pandita, "Music Genre Classification using Random Forest - hackerdawn - Medium," Medium, Jan. 06, 2022.

Robust to Overfitting

Can handle unbalanced Data

Computational Costs and Discontinuous Transition between classes

PAPER 3:
MUSIC GENRE CLASSIFICATION
STEPHEN N. M., 2023


CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

Music Genre Classification

A thesis submitted in partial fulfilment of the requirements

for the degree of Master of Science

in Computer Science

By

Nithil Mariya Stephen

May 2023

The objective of this research is to develop a precise and effective music genre classification model using Convolutional Neural Networks (CNN), Support Vector Machines (SVM) and Random Forest algorithms.

Kaggle Dataset: GTZAN Dataset - Music Genre Classification

- It consists of 1000 audio files, each 30 seconds long, from ten different music genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock
- It contains audio files in WAV format with a sample rate of 22050 Hz and a bit depth of 16 bits. The audio files were sampled from the Million Song Dataset and preprocessed to ensure high quality and the absence of irrelevant noise.

Reference: N. M. Stephen and California State University, Northridge, “Music Genre Classification,” thesis, 2023.

```
SVM Accuracy: 0.8988988988988988
SVM Precision: 0.8988440324863769
SVM Recall: 0.8985912972552196
SVM F1 Score: 0.8982088136367036
```

Figure 6: SVM Accuracy, Precision, Recall, F1score of the Model

```
Random Forest Classifier Accuracy: 0.8843843843843844
Random Forest Classifier Precision: 0.884493868939597
Random Forest Classifier Recall: 0.8845171727743688
Random Forest Classifier F1 Score: 0.8831142242214198
```

Figure 8: Random Forest Accuracy, Precision, Recall, F1 Score

```
63/63 [==============================] - 0s 4ms/step
Precision: 0.87
Recall: 0.87
F1 score: 0.87
```

```
score = cnn_model.evaluate(X_test_cnn, y_test, verbose=0)
print('CNN accuracy:', score[1])
```

```
CNN accuracy: 0.8678678870201111
```

Figure 4: CNN Accuracy, Precision, Recall, F1score of the Model

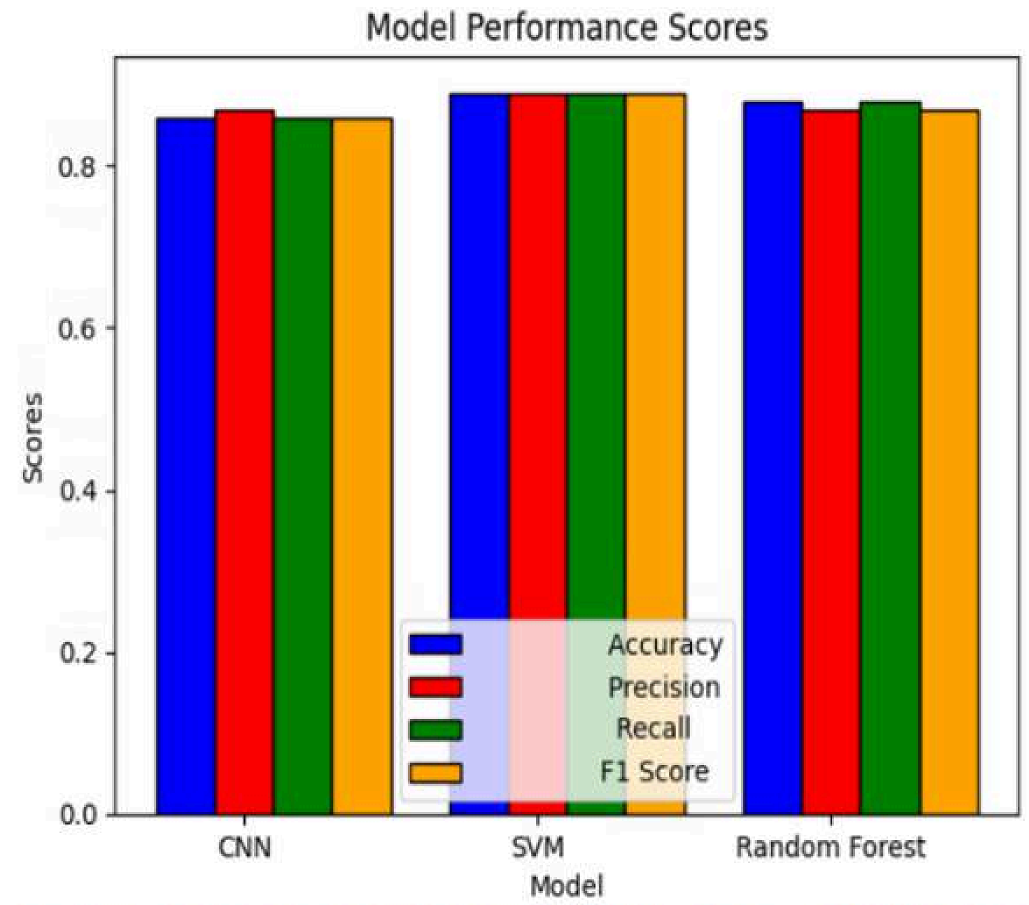


Figure 10: Comparison between CNN, SVM & Random Forest Model

Insights into accuracy and execution times of models

Detailed comparative analysis

Small size and narrow representation of genres

TARGETS

Ability to extract audio features and lyrics

Dataset usable for Bollywood Music

Genre AND Sentiment Analysis

DATASET AND FEATURES PREPROCESSING

SPOTIFY-TRACKS-DATASET V1 (@MAHARSHIPANDYA)

dataset.csv (20.12 MB)

Detail Compact Column 5 of 21 columns

	artists	album_name	track_name	track_genre
	str	str	str	str
ues	31438 unique values	46590 unique values	73609 unique values	114 unique values
g5xgaYa	Matrix & Futurebound;Luke Bingham	All I Know EP (feat. Luke Bingham)	All I Know - M&F's Rolling Out Radio Mix	drum-and-bass
Ma7JcJ2	Buren Van De Brandweer	Weekend Weg	Weekend Weg	hardstyle
M0gAmeI	Red Hot Chili Peppers	Return of the Dream Canteen	Reach Out	alt-rock
M0gAmeI	Red Hot Chili Peppers	Return of the Dream Canteen	Reach Out	funk
M0gAmeI	Red Hot Chili Peppers	Return of the Dream Canteen	Reach Out	metal
gsi8Ism	Jorge Drexler	Sus primeras grabaciones 1992-1994 (La luz que	No te creas	afrobeat

Reference: https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset/data

DATASET COLLECTION SPOTIPY

By carrying out EDA/model training, we found gaps in our chosen dataset

STEP 1

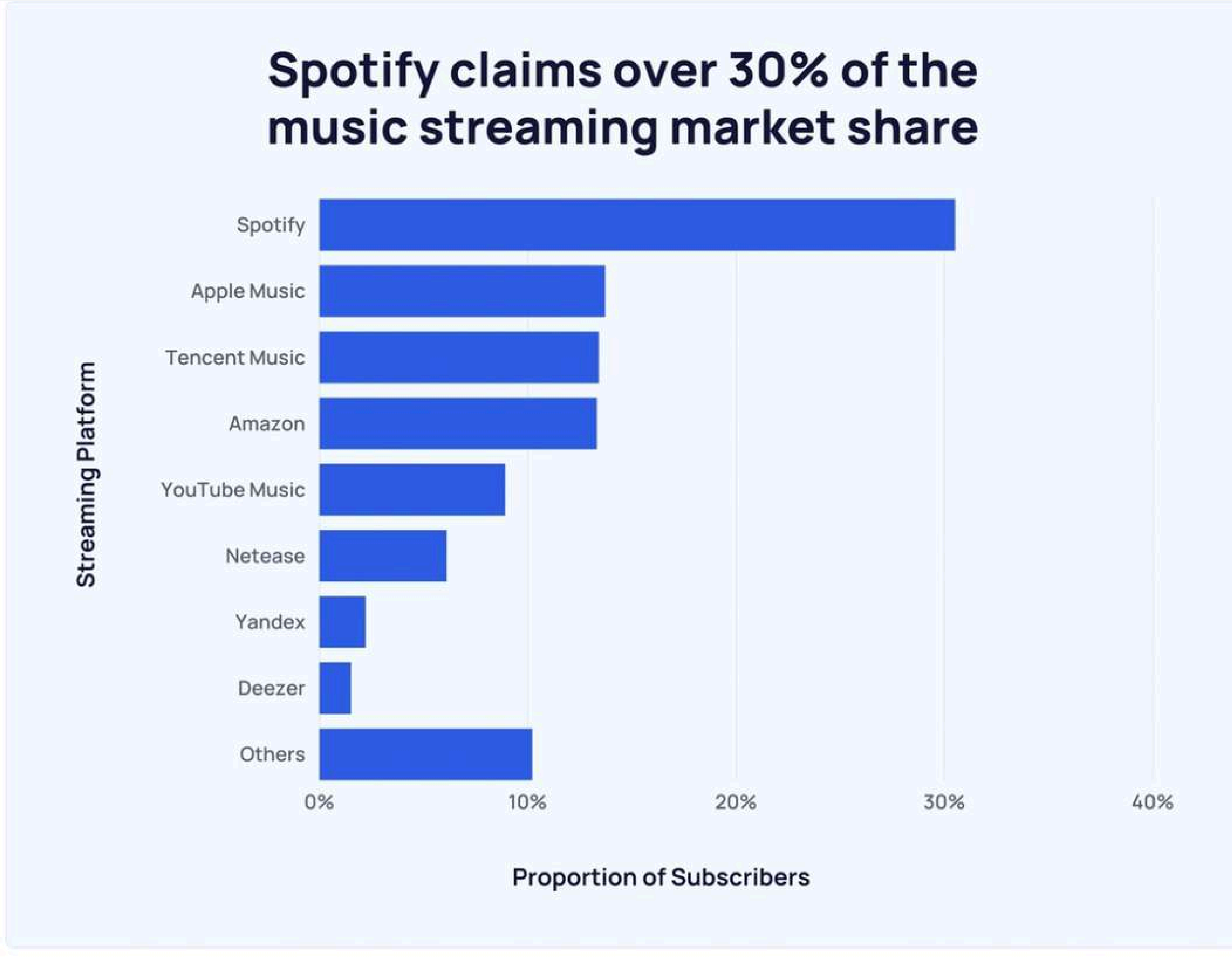
We populated our database with additional data using Spotify's API - **Spotipy**

STEP 2

We repeated this procedure for our model iterations - using Playlists already curated by Spotify to minimise Sentiment Bias

STEP 3

WHY SPOTIPY?



[Music Streaming Services Stats (2024)]

DATASET COLLECTION GENIUS API

Extract song and artist names from our curated database

STEP 1

Use the **Genius API** to get lyrics for these songs

STEP 2

Reference: Genius API

FEATURES IN OUR DATA

1. acousticness
2. loudness
3. danceability
4. energy
5. duration_ms
6. speechiness
7. valence
8. tempo
9. instrumentalness
10. liveness
11. unnamed
12. track_id
13. artists / artist_name
14. album_name
15. track_name
16. popularity
17. explicit
18. key
19. mode
20. time_signature
21. track_genre
22. type
23. id
24. uri
25. track_href
26. analysis_url
27. mood
28. lyrics

FEATURES EXTRACTED

1. **acousticness**: A confidence measure from 0.0 to 1.0 of whether the track is acoustic.
2. **loudness**: The overall loudness of a track in decibels (dB). Positive values represent louder songs while negative values suggest quieter ones.
3. **danceability**: Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable
4. **energy**: Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.
5. **duration_ms**: The track length in milliseconds.
6. **speechiness**: It detects the presence of spoken words in a track. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

FEATURES EXTRACTED

7. **valence**: A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive, while tracks with low valence sound more negative.
8. **tempo**: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
9. **instrumentalness**: Predicts whether a track contains no vocals. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content.
10. **liveness**: Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
11. **lyrics**: It has entire lyrics of the song (extracted using Genius API).

FEATURES PREPROCESSING
USING RANDOM FOREST

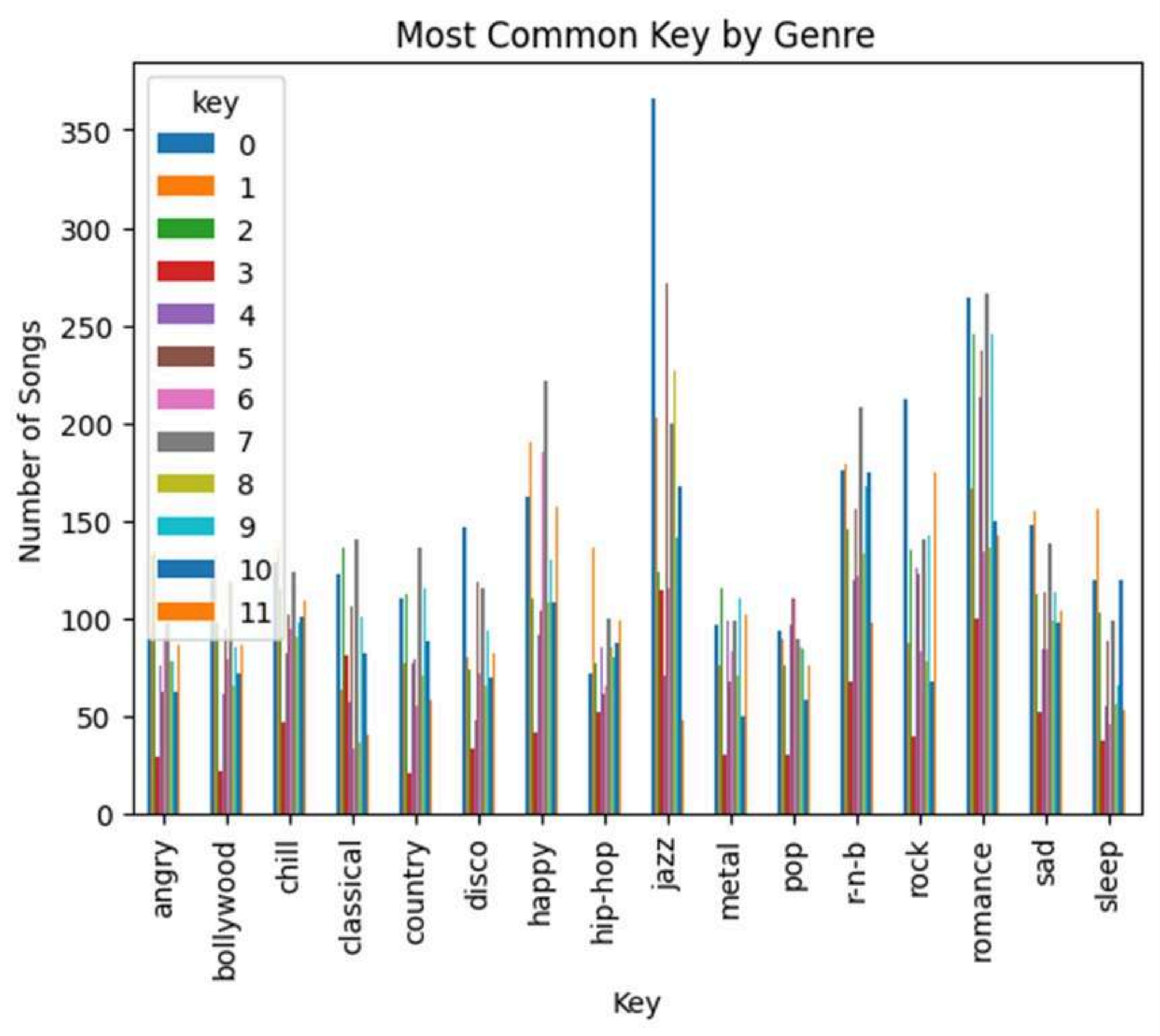
Sentiment

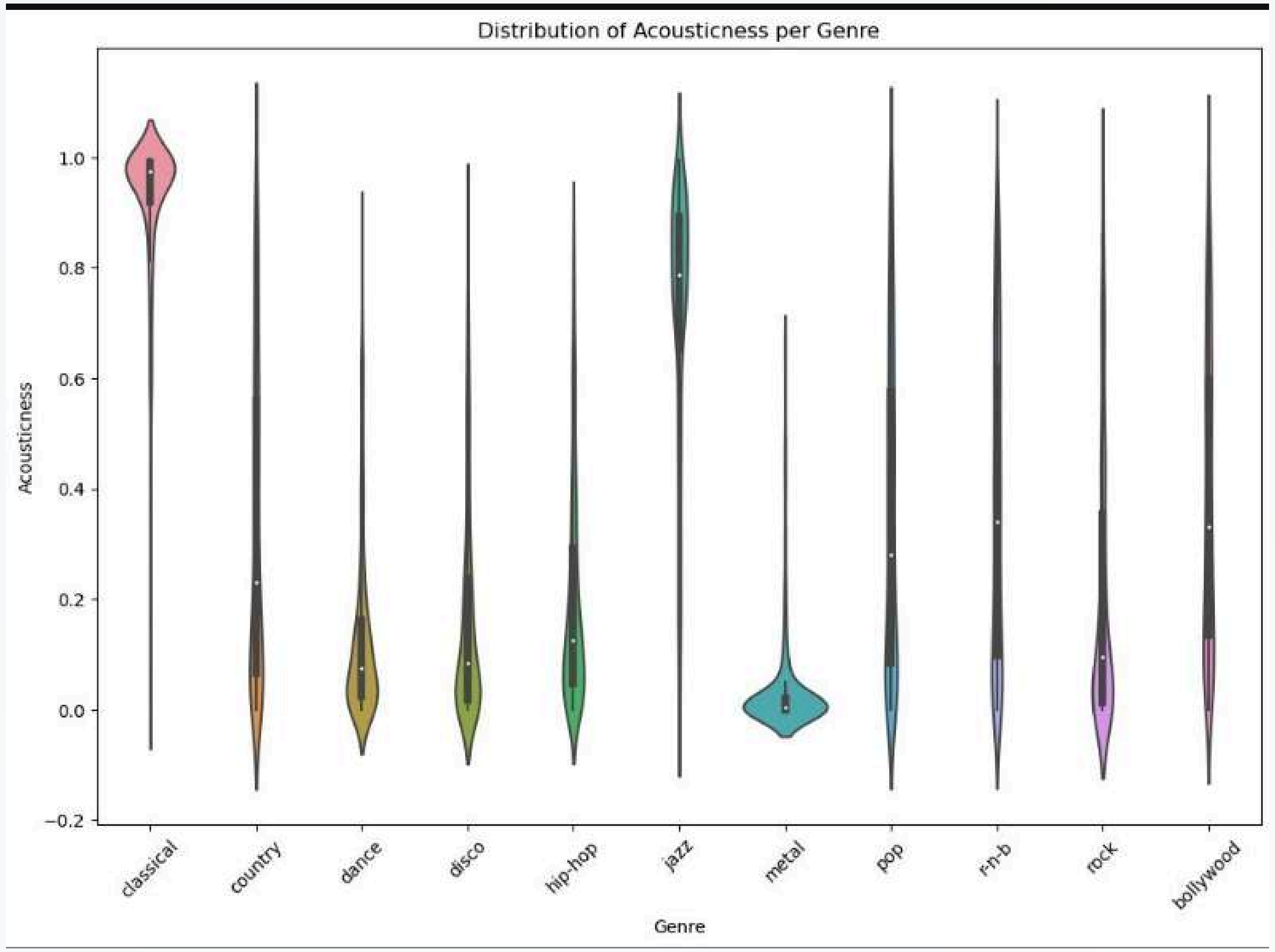
	feature	importance
1	danceability	0.150815
3	energy	0.134103
0	acousticness	0.129130
12	valence	0.106760
7	loudness	0.105663
4	instrumentalness	0.086692
10	tempo	0.069721
2	duration_ms	0.068481
9	speechiness	0.052696
6	liveness	0.044743
5	key	0.024582
11	time_signature	0.016400
8	mode	0.010211

Genre

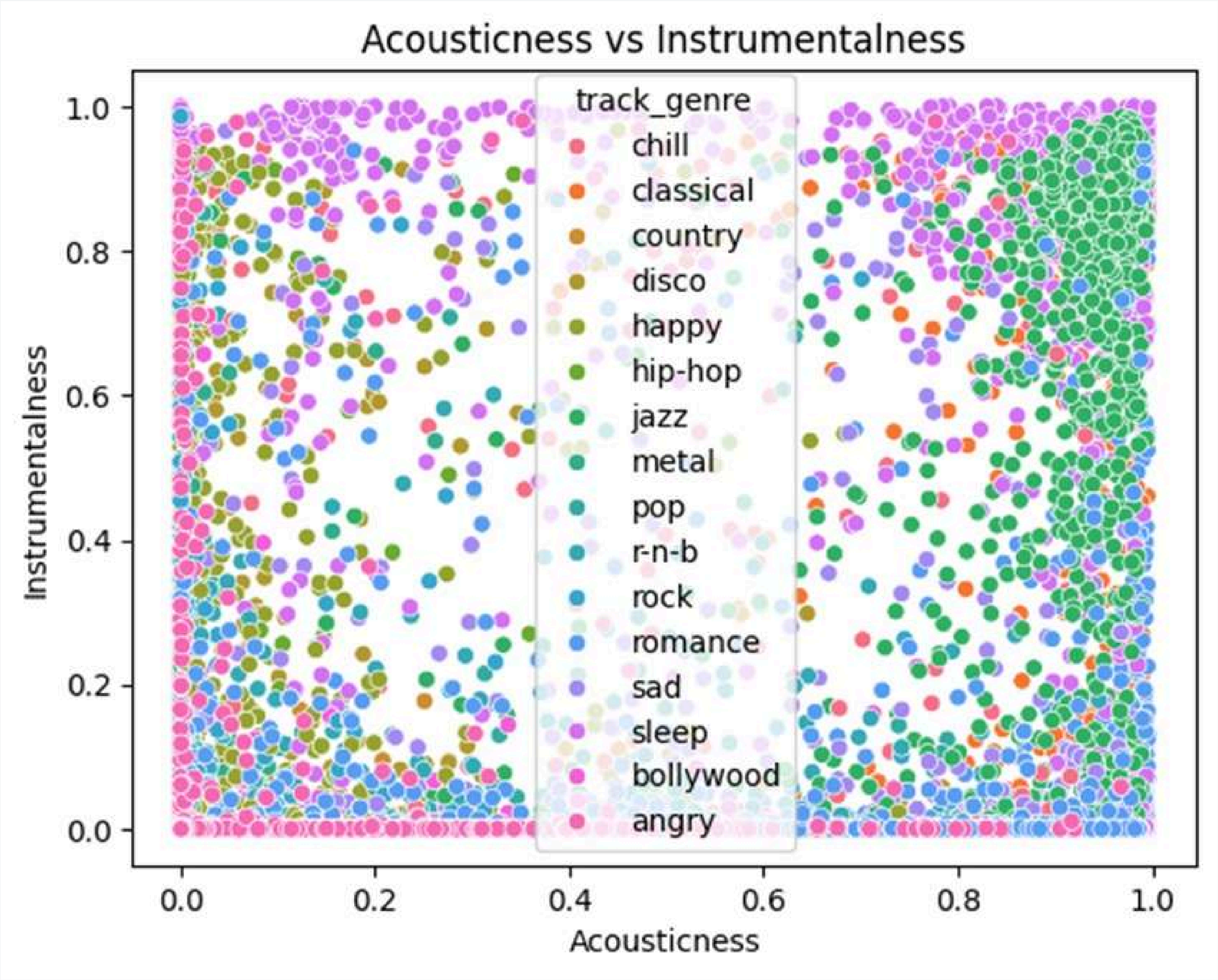
	feature	importance
0	acousticness	0.135235
2	duration_ms	0.110780
7	loudness	0.099557
1	danceability	0.098635
3	energy	0.098456
9	speechiness	0.085940
12	valence	0.082445
10	tempo	0.076304
4	instrumentalness	0.072420
6	liveness	0.071031
5	key	0.046992
8	mode	0.014148
11	time_signature	0.008056

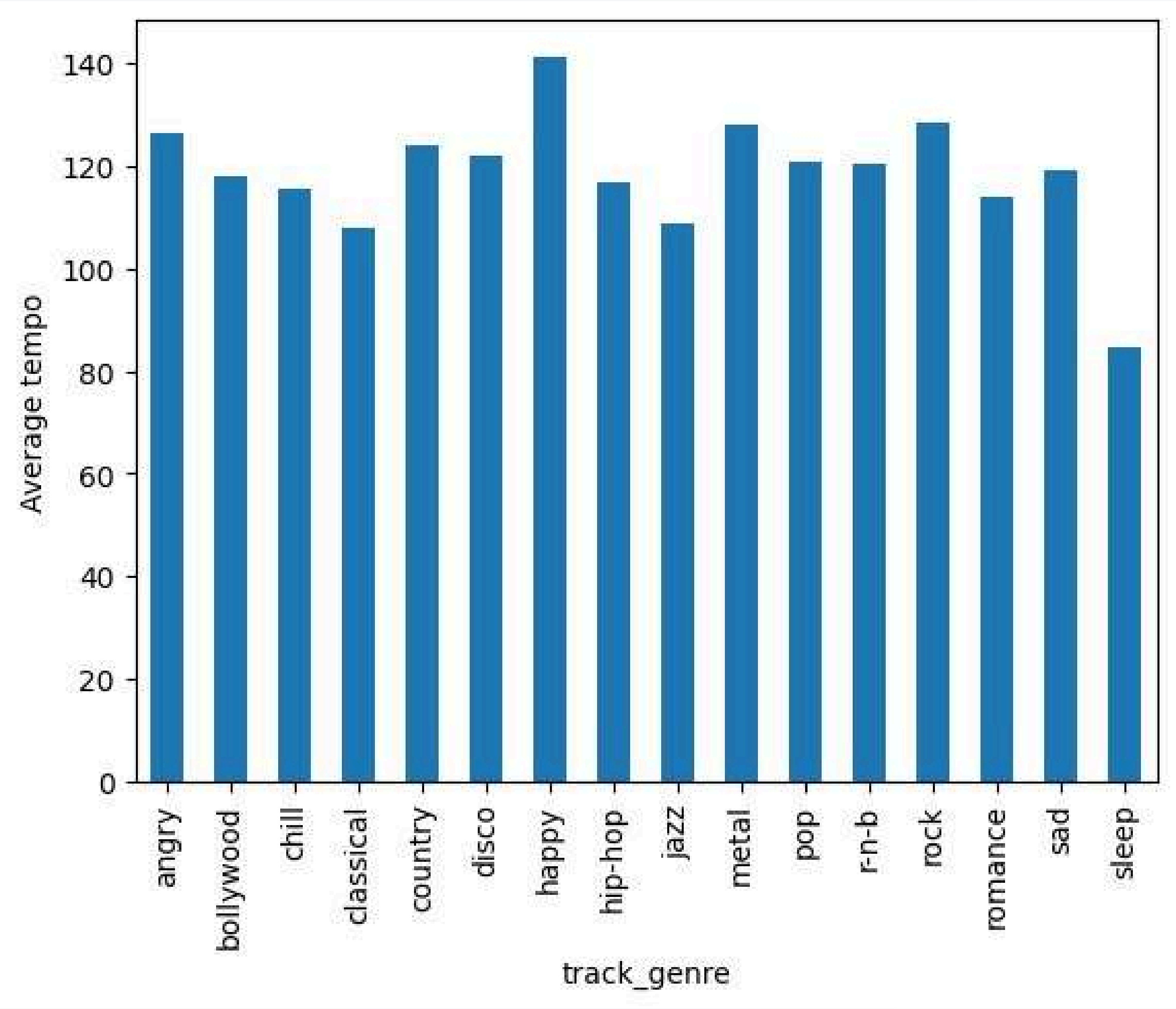
FEATURES PREPROCESSING



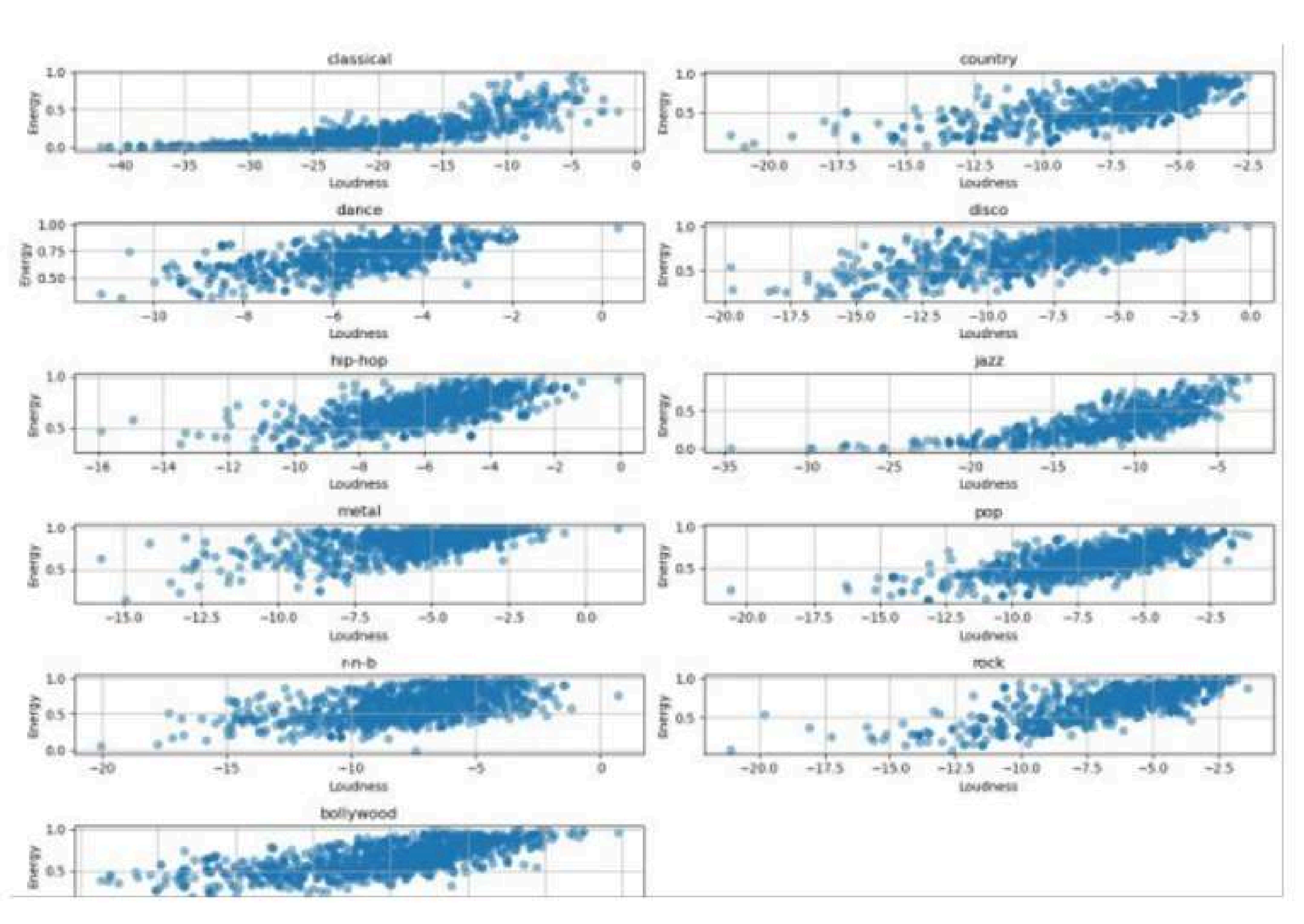


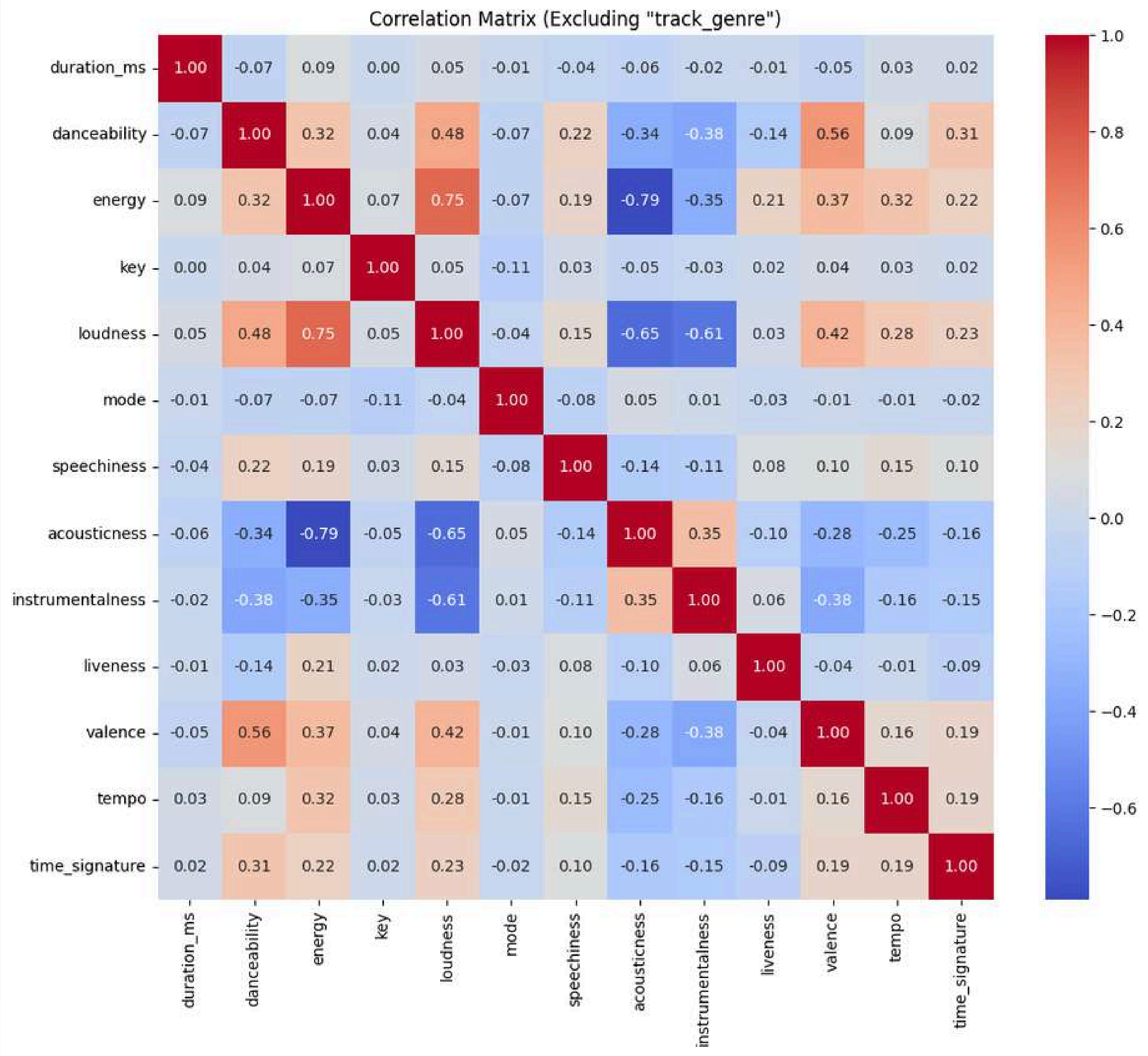
FEATURES PREPROCESSING



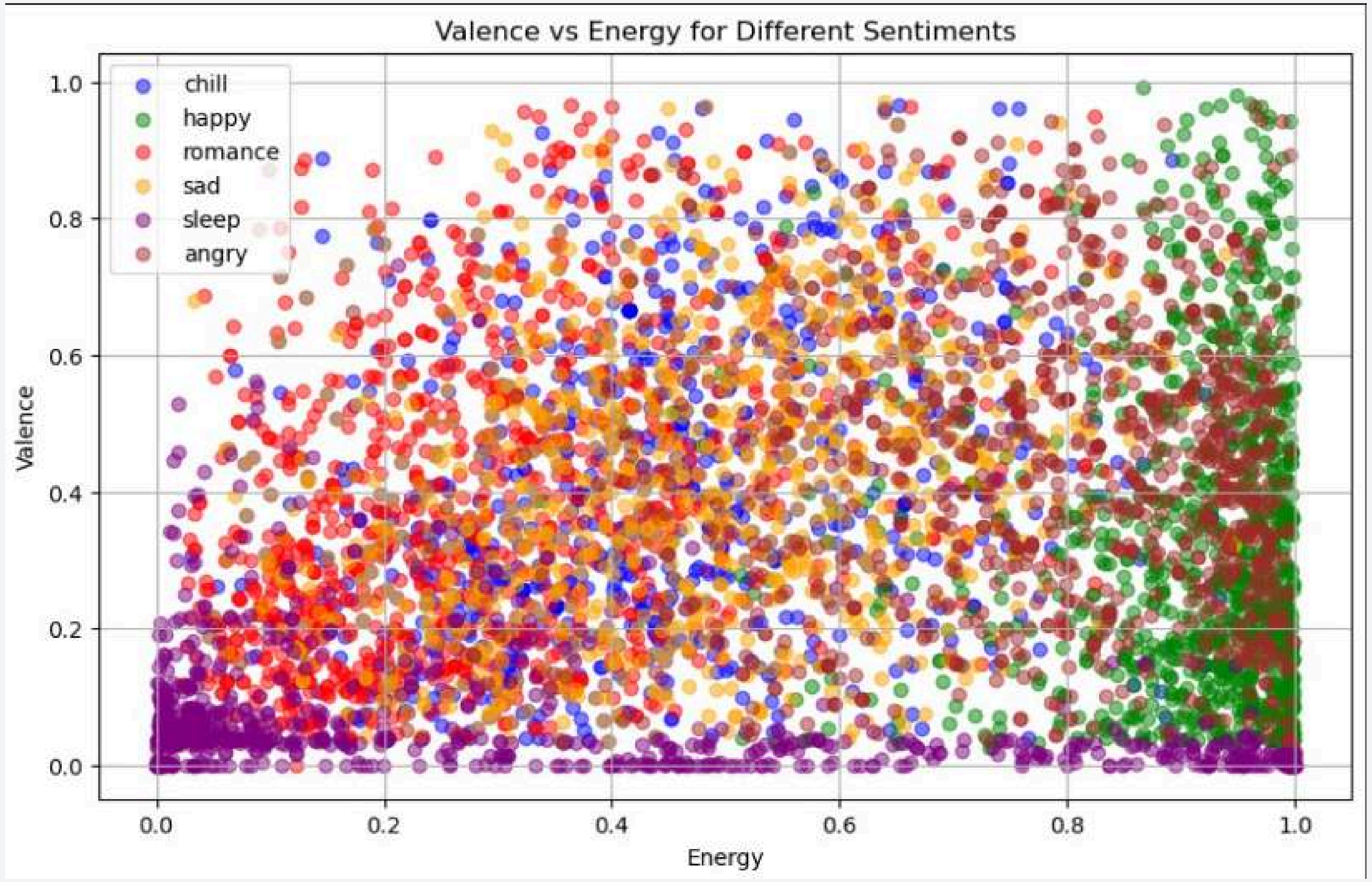


FEATURES PREPROCESSING





FEATURES PREPROCESSING



FEATURES PREPROCESSING LYRICS

```
Top 20 most common words in the lyrics:
one: 80315
like: 63558
would: 54296
de: 45670
know: 41509
said: 38021
time: 34331
see: 34074
man: 32696
us: 32100
could: 29509
love: 29132
never: 28042
go: 28000
day: 27823
feat: 26745
say: 26721
back: 24475
little: 24022
might: 23899
```

```
Summary Statistics for Number of Words:
count       6116.000000
mean        3761.653859
std        14889.676919
min            3.000000
25%          254.000000
50%          350.000000
75%          559.000000
max       149178.000000
Name: num_words, dtype: float64

Summary Statistics for Number of Characters:
count       6116.000000
mean       21011.215827
std        82822.459365
min           23.000000
25%         1324.000000
50%         1820.000000
75%         2908.750000
max       825594.000000
Name: num_characters, dtype: float64
```

```
Summary Statistics for Average Word Length:
count    6116.000000
mean        5.119701
std         0.645168
min         2.464286
25%         4.717478
50%         5.056374
75%         5.400000
max        12.125000
Name: average_word_length, dtype: float64

Summary Statistics for Vocabulary Richness:
count    6116.000000
mean        0.509085
std         0.171208
min         0.062331
25%         0.397610
50%         0.495708
75%         0.602236
max         1.000000
Name: vocabulary_richness, dtype: float64

Summary Statistics for Readability Score (Flesch Reading Ease):
count    6116.000000
mean      -44.441696
std       172.666224
min     -4351.890000
25%      -111.162500
50%         9.260000
75%        60.350000
max       110.260000
Name: readability_score, dtype: float64
```

ML METHODOLOGY

01

02

03

FEATURE EXTRACTION

Extract Song Audio Features and Lyrics by taking Track Name and Artist as Inputs

RANDOM FOREST

3 Forests:
Audio Features for Genre
Audio Features for Sentiment
Lyrics for Sentiment

RESULTS

Our Output is the Genre and Sentiment!

01 POPULATING DATASET

Added extra songs for sentiments / genres and lyrics as per imbalances in data

mood	
chill	877
sad	690
sleep	587
happy	586
romance	105

02 RANDOM FOREST

Fine-Tuned min_leves, max_depth and n_estimators for all trees using ROC-AUC Curve

03 EVALUATION

using standardised testing methods to assess model outputs on test and train datasets (We used an 20-80 split)

mood	
romance	1326
happy	1183
chill	1099
sad	988
angry	933
sleep	587

RANDOM FOREST

Classification of Music Genres using Feature Selection and Hyperparameter Tuning

August 2022 · Journal of Artificial Intelligence and Capsule Networks 4(3):167-178

DOI:10.36548/jaicn.2022.3.003

Authors:

Rahul Singhal
New York University

Random Forests are known for their ability to resist overfitting, a common problem where the model performs well on training data but poorly on unseen data. This is because they ensemble multiple decision trees, each trained on a random subset of features and data points. This inherent randomness reduces the variance of the model, leading to better generalization

Model	Accuracy (Test data)	F1- Score (Test data)	ROC Auc Score (Test data)	Accuracy (Validation data)	F1 Score (Validation data)	Accuracy (Training data)	F1 Score (Training data)
Logistic Regression	51.19	49.86	88.66	50.51	49.7	50.87	50.04
KNN	56.75	56.54	92.73	55.56	55.46	60.39	60.36
SVM	54.06	53.65	90.81	54.11	53.48	54.41	53.66
XGBoost	**99.60**	**99.60**	**99.99**	**99.74**	**99.74**	**84.66**	**84.68**
Random Forest	**99.60**	**99.60**	**99.71**	**99.72**	**99.72**	**97.2**	**97.2**

Results-of-different-Machine-learning-models-on-all-features

SUPPORT VECTOR MACHINE

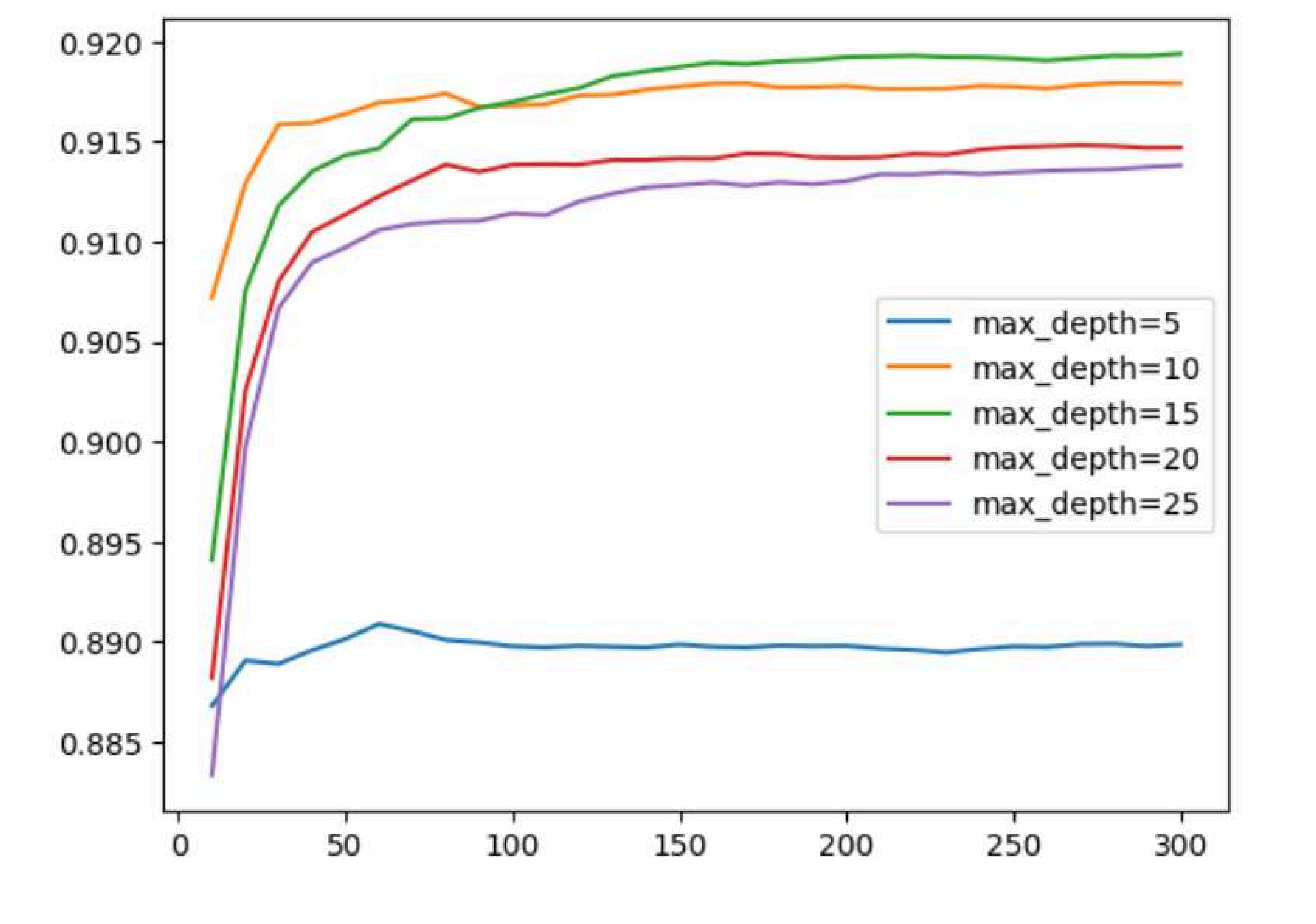
SVMs offer a robust approach to music genre and sentiment classification with their ability to handle high-dimensional data, achieve good generalization, and tackle non-linear relationships. Additionally, interpretability techniques can provide valuable insights into the music characteristics driving the classifications.

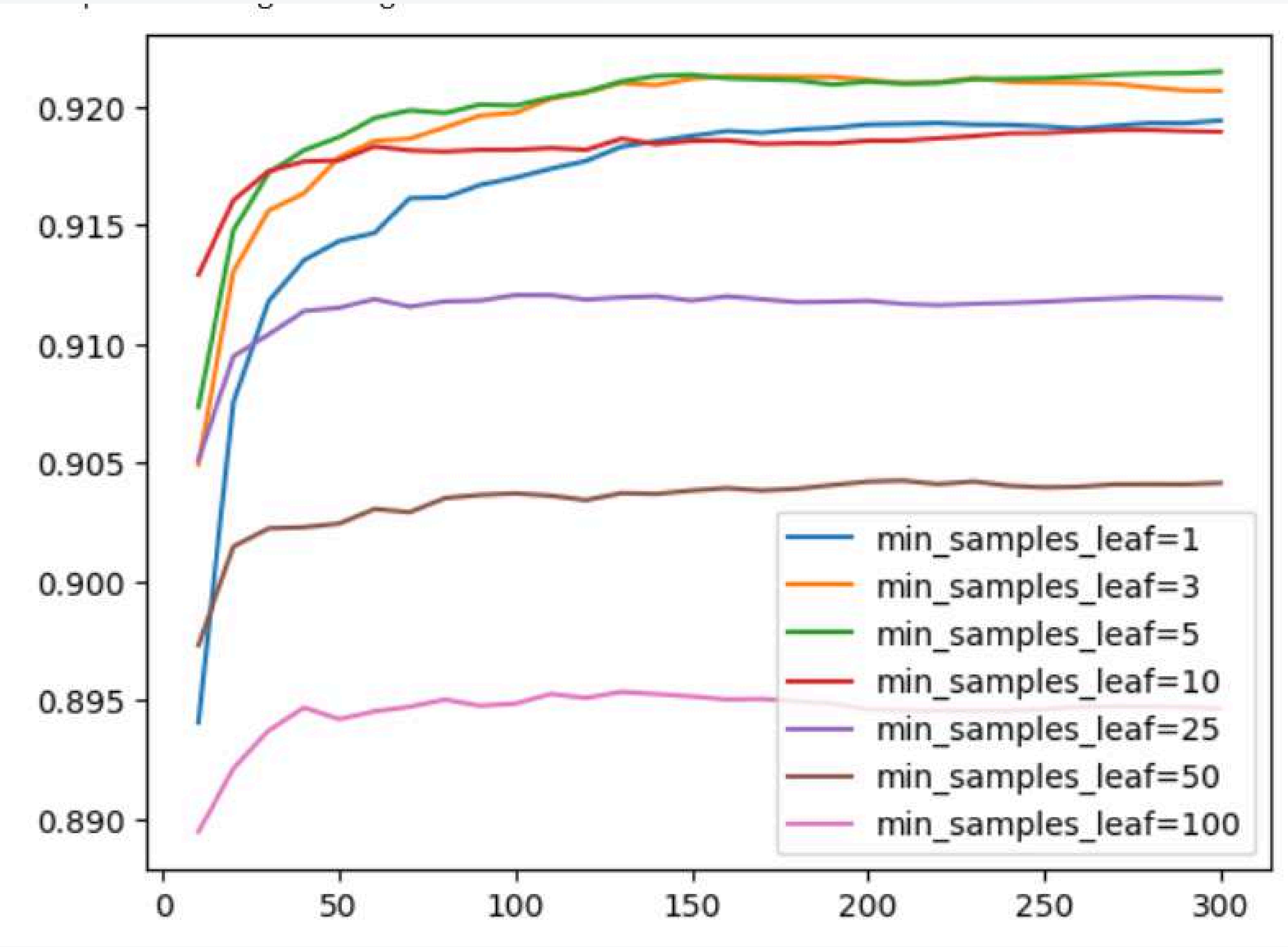
Classifier	Lenient Accuracy	%
Decision Tree	0.7491782553729498	74.91%
SVM	0.8165865992414686	81.65%
K-Nearest Neighbours	0.7173198482932994	71.73%
Gaussian Naïve Bayes'	0.7928445006321114	79.28%

Reference: T. Tismo-Capili, "Musical Emotions Analysis," thesis, 2021.

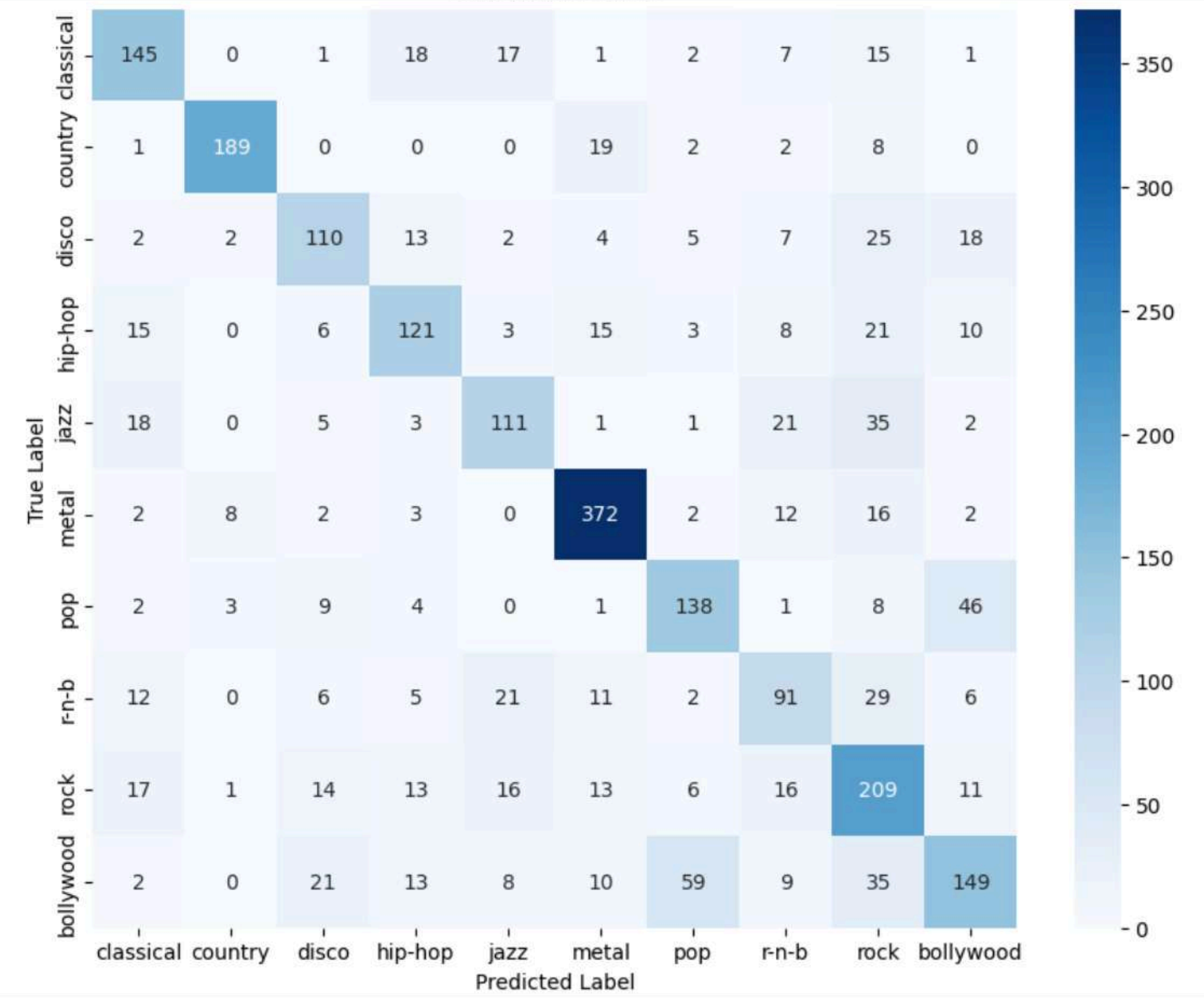
PERFORMANCE METRICS

FINE TUNING RANDOM FOREST USING ROC-AUC CURVE





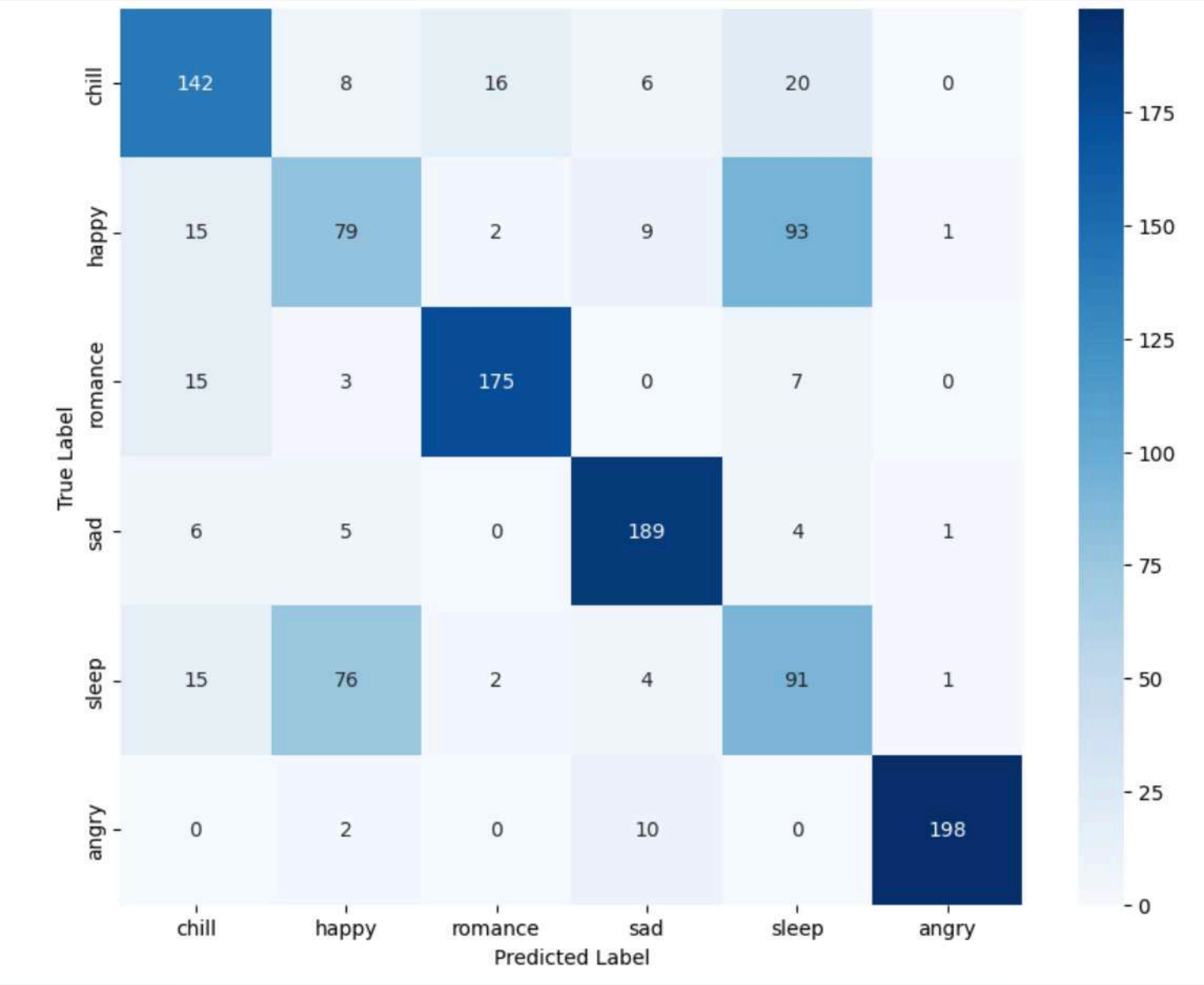
```
ROC AUC Score on Test Set: 0.9996658298379364
Accuracy: 0.6670746634026927
Precision: 0.6696210710924082
Recall: 0.6670746634026927
F1-score: 0.6658519165553969
```



GENRE CLASSIFICATION USING RF ON AUDIO FEATURES AND RE-POPULATED DATA

SENTIMENT CLASSIFICATION USING AUDIO FEATURES

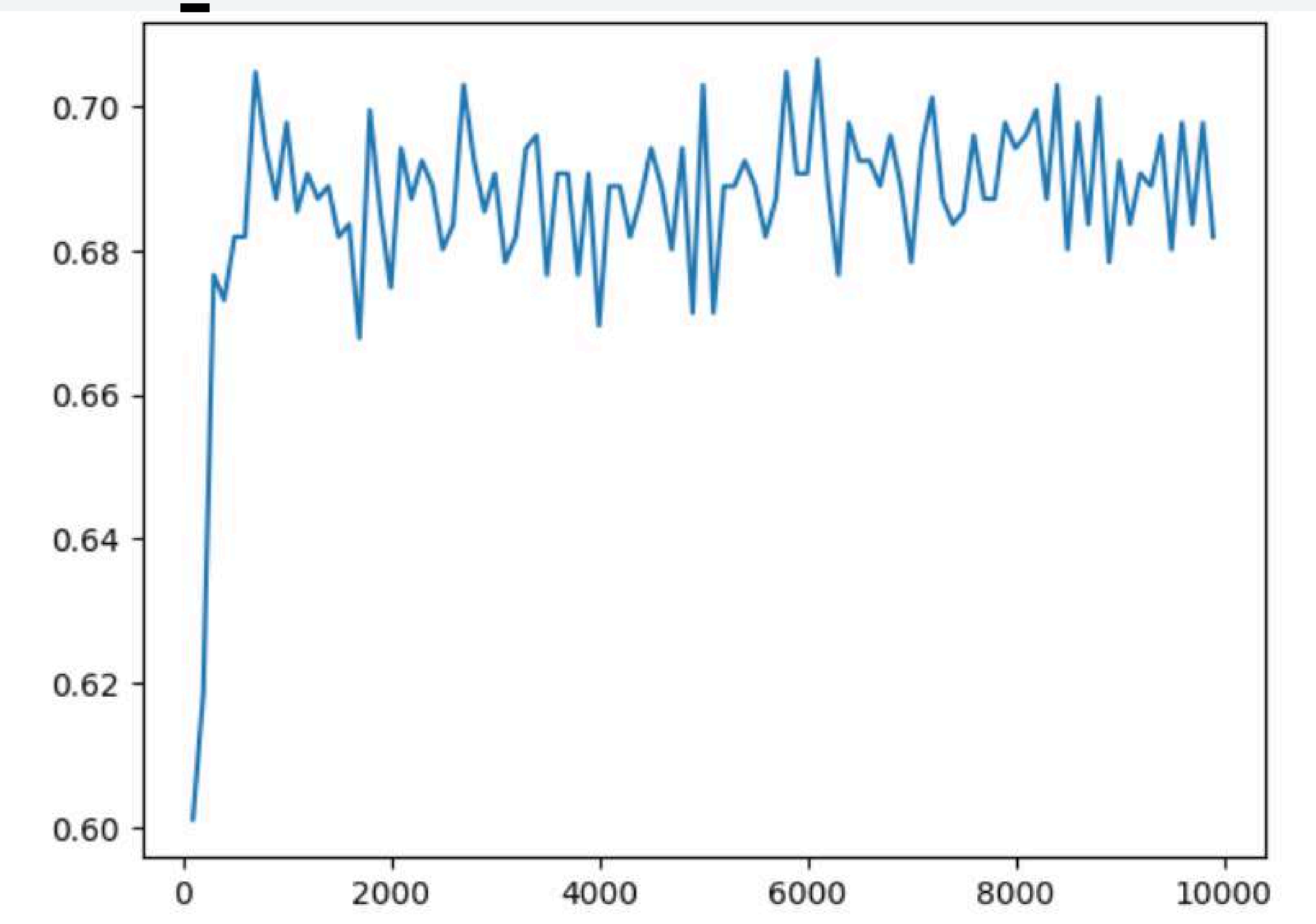
ROC AUC Score on Test Set: 0.9478372559061511
Accuracy: 0.7313807531380753
Precision: 0.7332340073966114
Recall: 0.7313807531380753
F1-score: 0.731413124872323



```
Number of components to capture 95% variance: 411
```

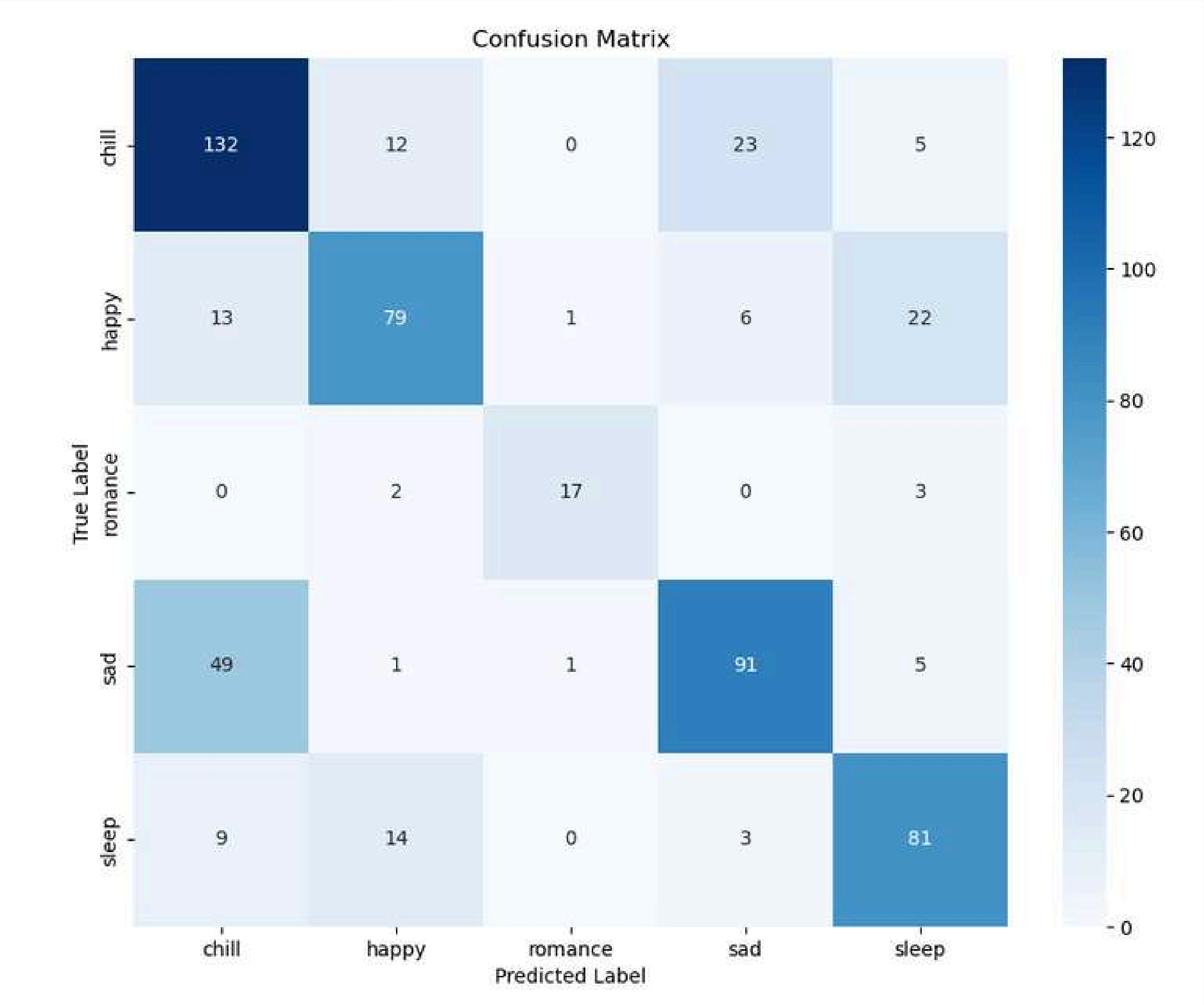
PCA FOR TEXT DATA

MAX_FEATURES V/S ACCURACY



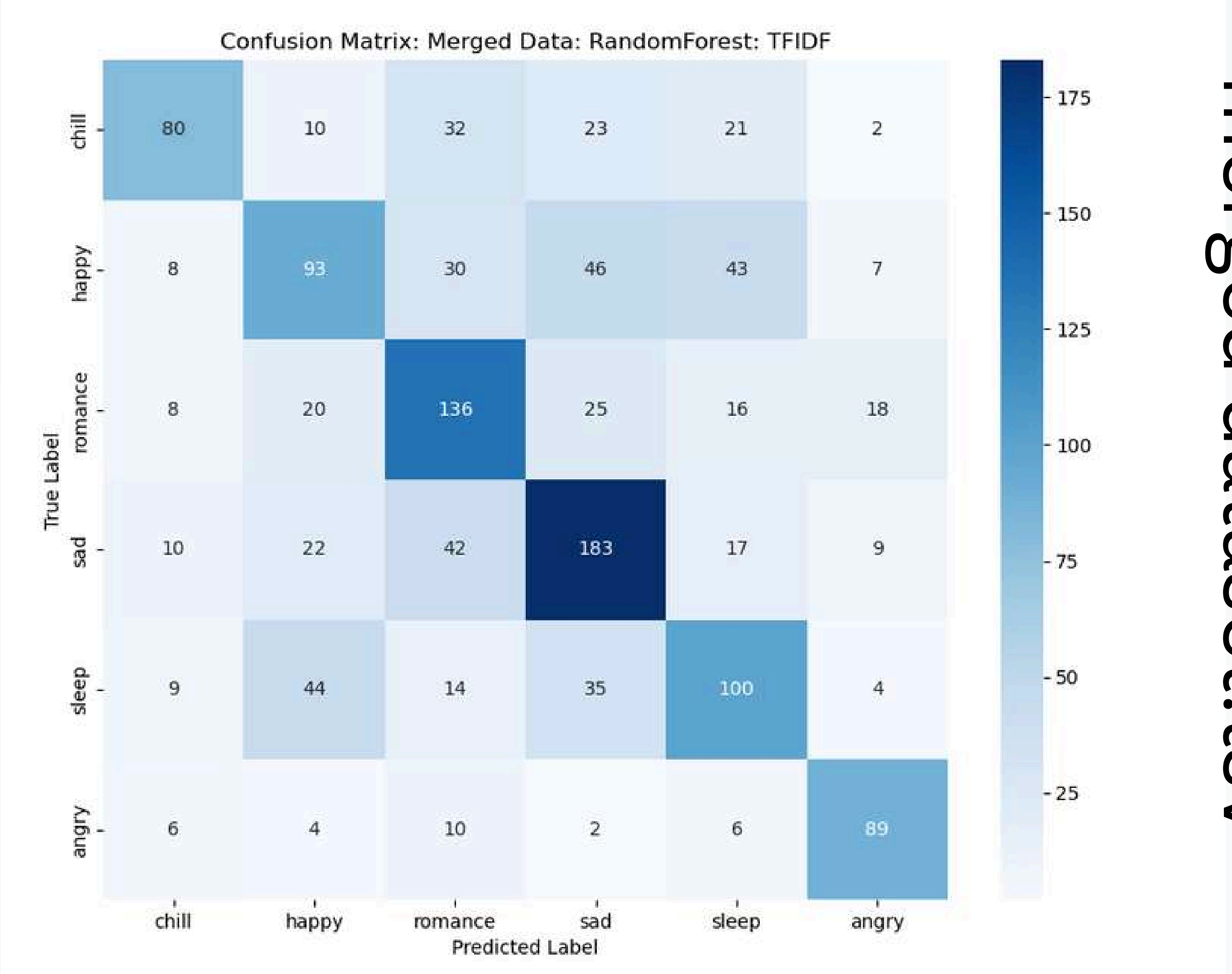
RANDOM FOREST + TFIDFVECTORIZER FOR SENTIMENTS

Unique_songs.tsv



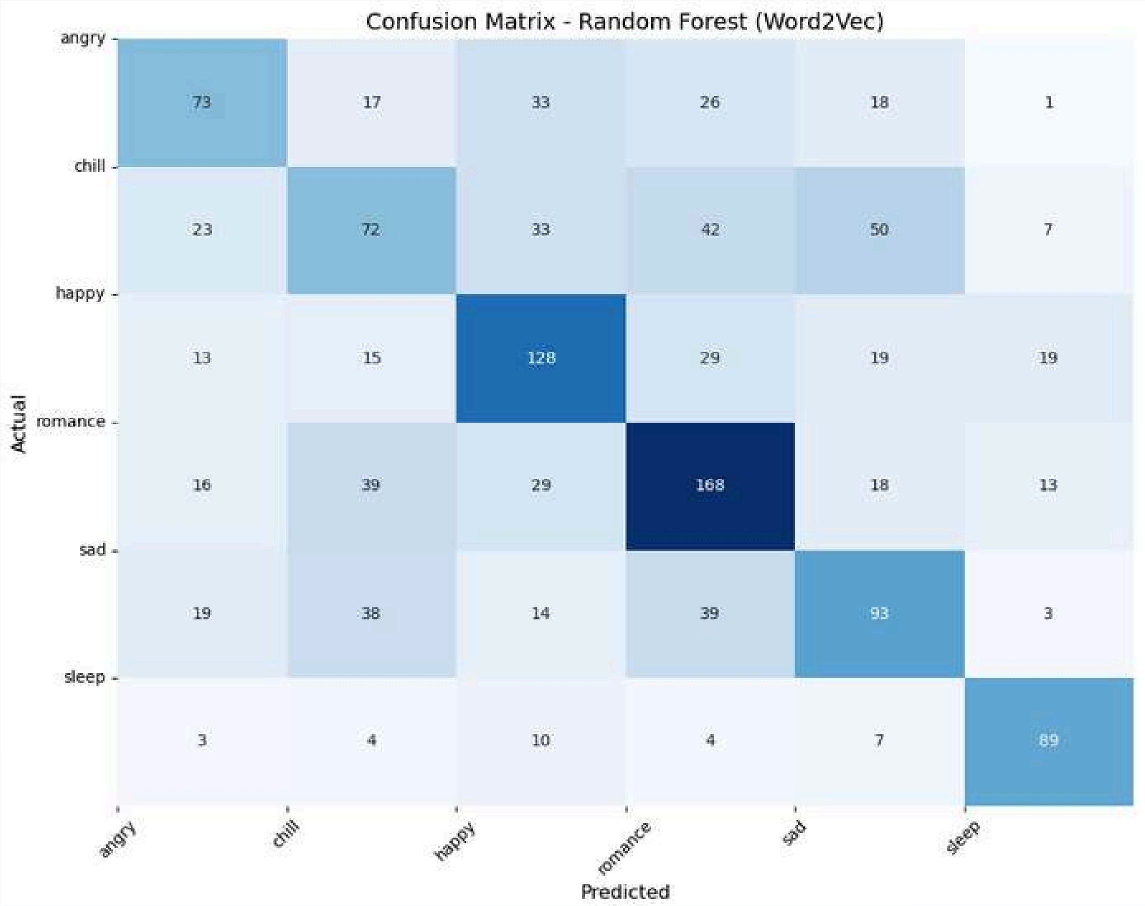
Overall Accuracy: 0.70298769771529
Precision: 0.7091519316695074
Recall: 0.70298769771529
F1 Score: 0.7023487683464928
ROC-AUC Score: 0.8996128398867761

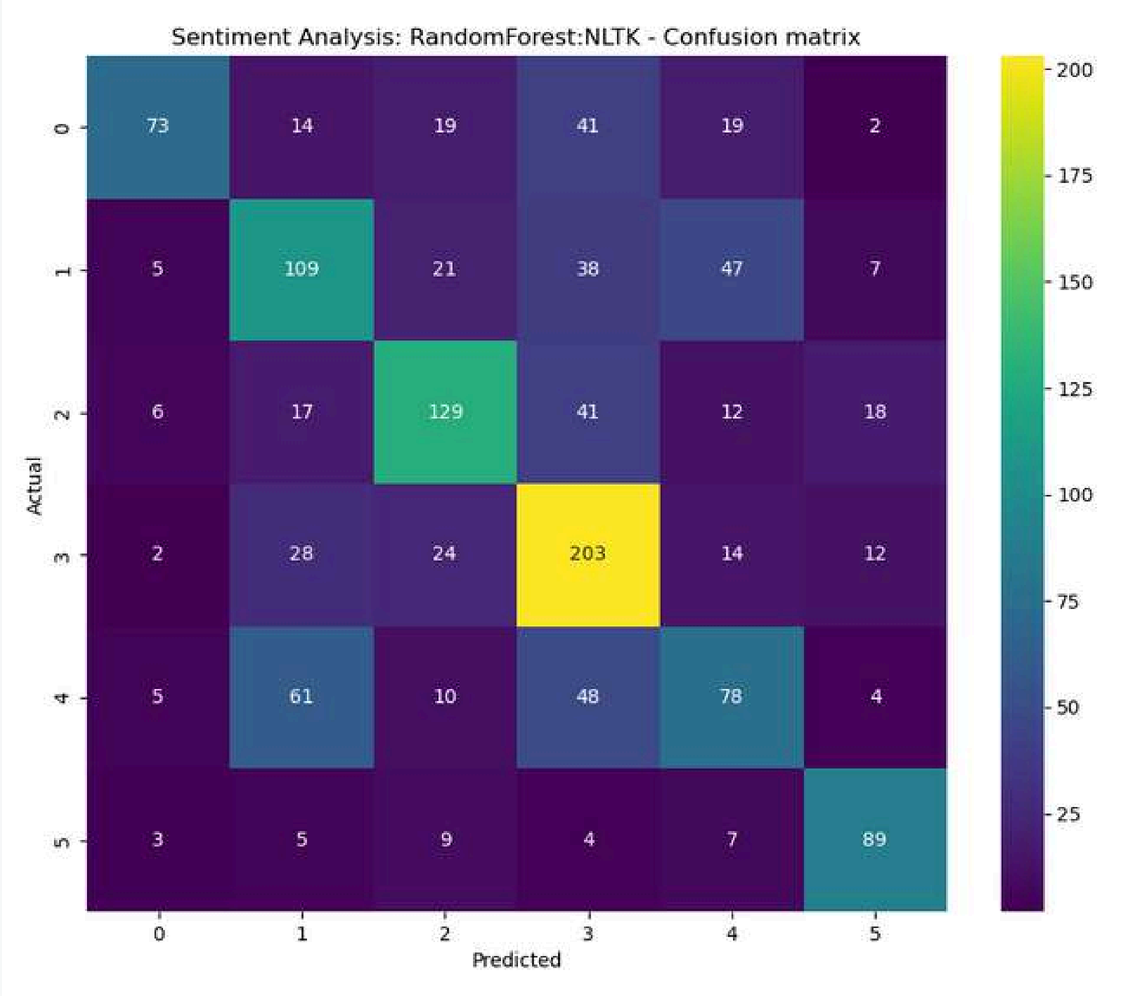
merged dataset.tsv



Overall Dataset Accuracy: 0.955846279640229
Precision: 0.5575722806349495
Recall: 0.5563725490196079
F1 Score: 0.5530872389008925
ROC-AUC Score: 0.833304250486889

RANDOM FOREST MORE METHODS






```
Accuracy: 0.5089869281045751
Precision: 0.5016674256107353
Recall: 0.5089869281045751
F1 Score: 0.5036180840551489
ROC-AUC Score: 0.8052826274839534
```

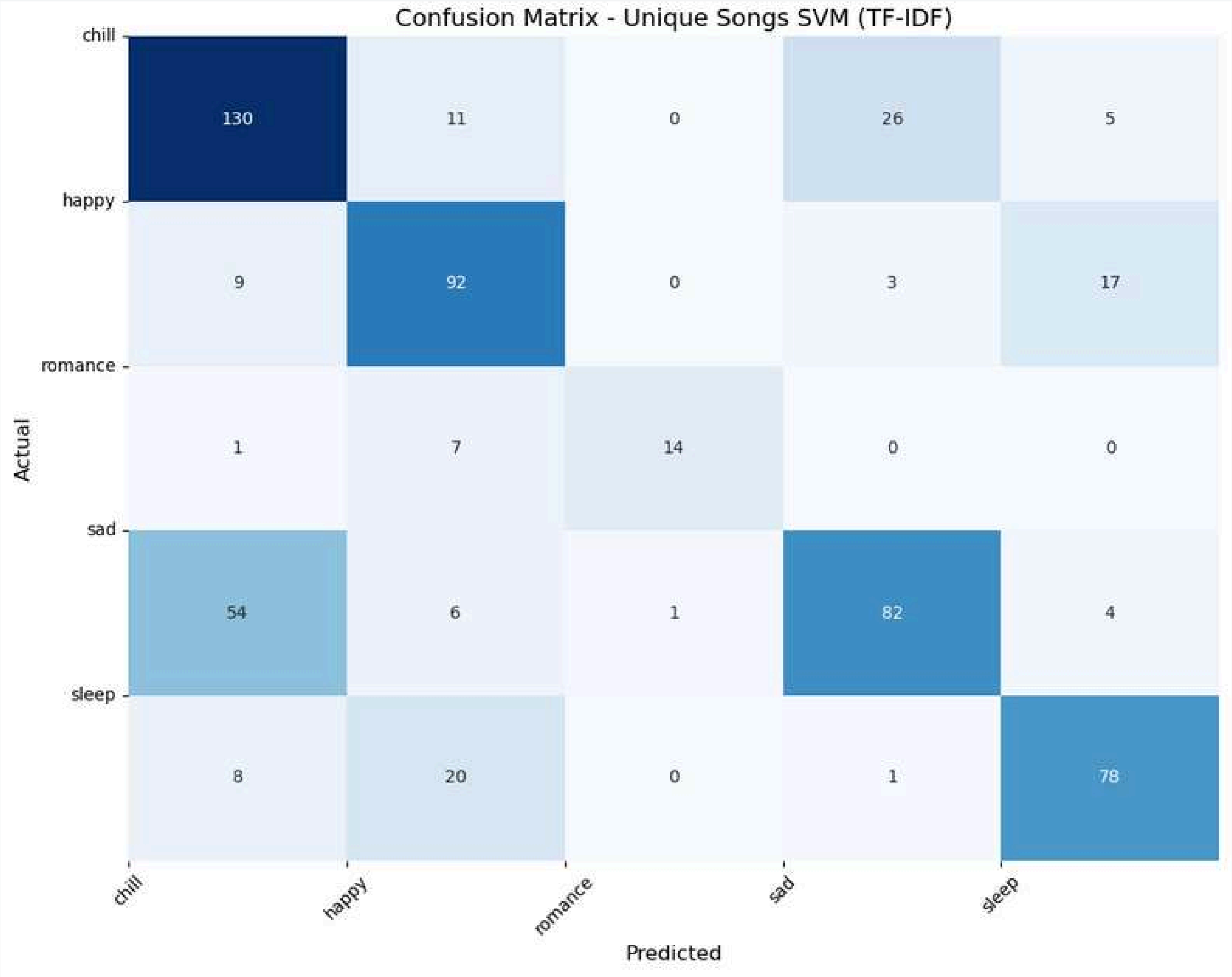
```
Accuracy: 0.5563725490196079
Classification Report:
              precision    recall  f1-score   support

       angry       0.78      0.43      0.56       168
       chill       0.47      0.48      0.47       227
       happy       0.61      0.58      0.59       223
     romance       0.54      0.72      0.62       283
         sad       0.44      0.38      0.41       206
       sleep       0.67      0.76      0.71       117

    accuracy                           0.56      1224
   macro avg       0.58      0.56      0.56      1224
weighted avg       0.57      0.56      0.55      1224
```

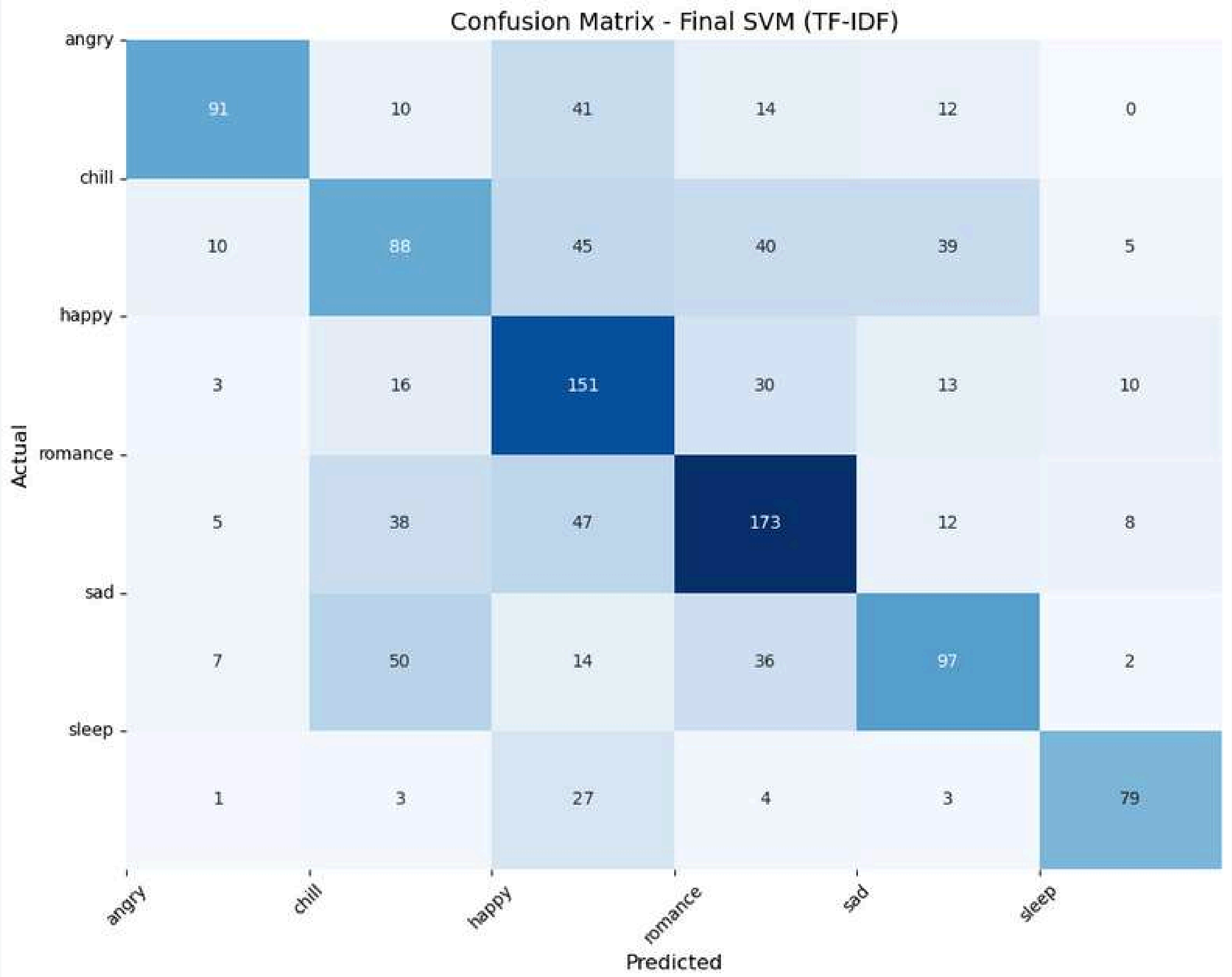
SUPPORT VECTOR MACHINE

Unique_songs.tsv



Accuracy: 0.6959578207381371
Recall: 0.6959578207381371
Precision: 0.7046649276216778
F1 Score: 0.6942728948628958
ROC-AUC Score: 0.9142420969525451

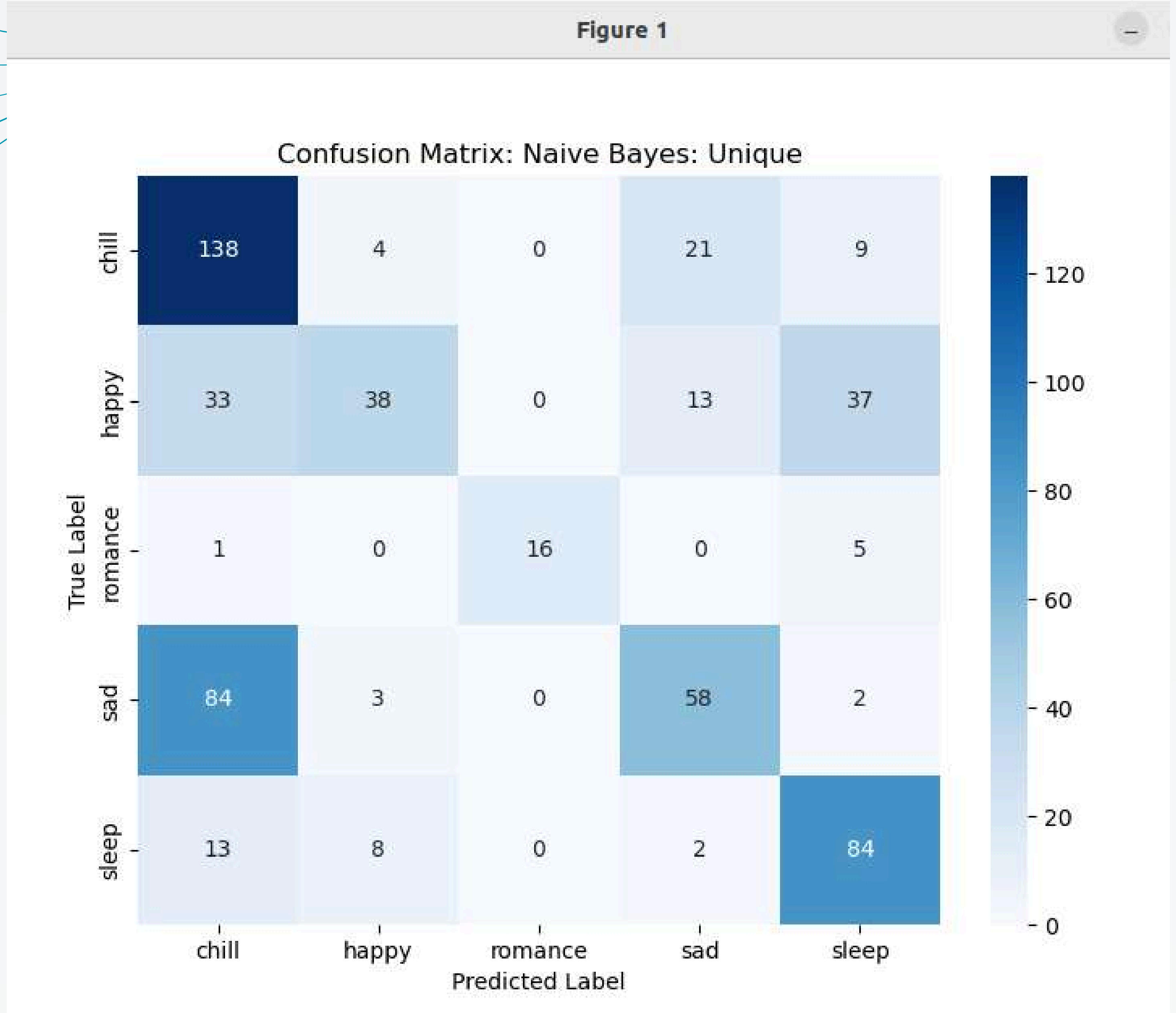
merged dataset.tsv



Accuracy: 0.5547385620915033
Recall: 0.5547385620915033
Precision: 0.5710571965924269
F1 Score: 0.5553503679684829
ROC-AUC Score: 0.8471514906598979

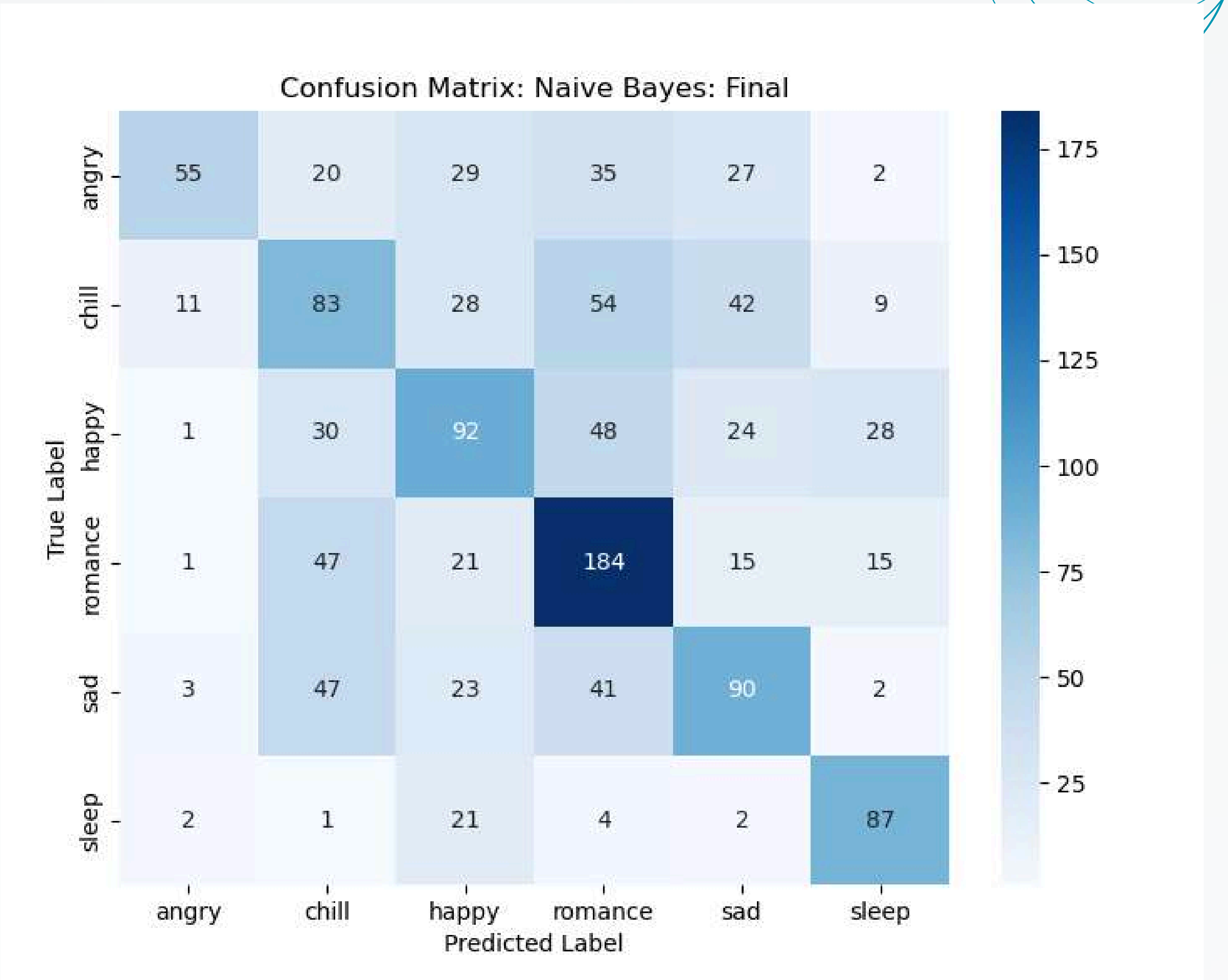
NAIVE BAYES

Unique_songs.tsv




```
Test Accuracy: 0.5869947275922671
Test Precision: 0.6209149424822115
Test Recall: 0.5869947275922671
Test F1 Score: 0.5684542475835702
```

merged dataset.tsv




```
Test Accuracy: 0.48284313725490197
Test Precision: 0.49937541542653013
Test Recall: 0.48284313725490197
Test F1 Score: 0.4767085615923789
```

CHALLENGES

Imbalanced dataset: We extracted same number of songs for each sentiment but Genius API did not have lyrics for all the songs, this will be an issue for multilingual sentiment analysis.

Ambiguity: Analyzing music sentiment and genre is complex due to varied perceptions and evolving genre definitions. Determining the importance of sentiments and genres is subjective, influenced by personal preferences.

Deployment: It is limited to one language, resources to work with Hindi music were not available. Songs get mislabeled due to conflicting lyrical and musical features, which further gets distorted due to sarcasm and changing language context over time.

THANK YOU

OUR TEAM

Manjree
Kothari

Roma
Sahu

Vandita
Lodha